Joint Spatial Division and Multiplexing

Ansuman Adhikary[†], Junyoung Nam[*], Jae-Young Ahn[*], and Giuseppe Caire[†]

Abstract

We propose Joint Spatial Division and Multiplexing (JSDM), an approach to multiuser MIMO downlink that exploits the structure of the correlation of the channel vectors in order to allow for a large number of antennas at the base station while requiring reduced-dimensional Channel State Information at the Transmitter (CSIT). This allows for significant savings both in the downlink training and in the CSIT feedback from the user terminals to the base station, thus making the use of a large number of base station antennas potentially suitable also for Frequency Division Duplexing (FDD) systems, for which uplink/downlink channel reciprocity cannot be exploited. JSDM forms the multiuser MIMO downlink precoder by concatenating a *pre-beamforming* matrix, which depends only on the channel second-order statistics, with a classical multiuser precoder, based on the instantaneous knowledge of the resulting reduced dimensional "effective" channels. We prove a simple condition under which JSDM incurs no loss of optimality with respect to the full CSIT case. For linear uniformly spaced arrays, we show that such condition is closely approached when the number of antennas is large. For this case, we use Szego's asymptotic theory of large Toeplitz matrices to design a DFT-based pre-beamforming scheme requiring only coarse information about the users angles of arrival and angular spread. Finally, we extend these ideas to the case of a two-dimensional base station antenna array, with 3-dimensional beamforming, including multiple beams in the elevation angle direction. We provide guidelines for the pre-beamforming optimization and calculate the system spectral efficiency under proportional fairness and max-min fairness criteria, showing extremely attractive performance. Our numerical results are obtained via an asymptotic random matrix theory tool known as "deterministic equivalent" approximation, which allows to avoid lengthy Monte Carlo simulations and provide accurate results for realistic (finite) number of antennas and users.

Keywords: Multiuser MIMO Downlink, Antenna Correlation, 3D Beamforming, Deterministic Equivalents.

[*] Mobile Communications Division, Electronics Telecommunications Research Institute, Daejeon, Korea.

[†] Ming-Hsieh Department of Electrical Engineering, University of Southern California, CA.

This work was supported by the IT R&D program of MKE/KEIT in Korea [Development of beyond 4G technologies for smart mobile services].

I. INTRODUCTION

In a Multiuser MIMO (MU-MIMO) downlink where a base station (BS) with M antennas serves K single-antenna user terminals (UTs) on the same time-frequency slot, and the channel fading coefficients can be considered constant over coherence blocks of T channel uses, [1] the high-SNR system spectral efficiency behaves at best as $M^\star(1-M^\star/T)\log\mathsf{SNR}+O(1)$, where $M^\star=\min\{M,K,T/2\}$. The upper bound yielding this behavior is obtained by letting all UTs cooperate and using the result of [1] on the high-SNR capacity of the non-coherent block-fading MIMO point-to-point channel. A tight lower bound is obtained by devoting $M^\star$ dimensions per block to training, in order to acquire the Channel State Information at the Transmitter (CSIT), i.e., to estimate the downlink channel matrix on each fading coherence block. In Frequency Division Duplexing (FDD) systems, where the fading channel reciprocity cannot be exploited, the lower bound is achievable by assuming ideal instantaneous CSIT feedback from the UTs to the BS, between the downlink training phase and the data transmission phase. If, more realistically, instantaneous feedback in the same fading coherence block is not possible, a prediction error further decreases the system multiplexing gain by the factor $\max\{1-2B_dT_s,0\}$, where $B_d=vf_0/c$ is the Doppler bandwidth (Hz), (v denoting the UT speed (m/s), f_0 the carrier frequency (Hz) and c the light speed (m/s)), and T_s is the slot duration (s) [2], [3].

It is evident that, even not taking into account the cost of CSIT feedback (which may impact the uplink system capacity), the MU-MIMO multiplexing gain for an FDD system based on downlink training, channel estimation (and possibly prediction) at the UTs, and CSIT feedback, is significantly reduced when $M^\star$ is not much smaller than T and/or $2B_dT_s$ is not much smaller than 1. In particular, for large M and K, the downlink training represent a significant bottleneck (as quantified by the analysis in [4]) and the corresponding CSIT feedback yields an unacceptably high overhead for the uplink.

Alternatives that do not require CSIT [5] or require only outdated CSIT [6] (without requiring a strict one-slot prediction constraint) have been proposed. Although these schemes may achieve better multiplexing gain than the basic training and feedback scheme in certain conditions (see for example the comparison in [7]) they do not scale well with the number of BS antennas and UTs, since they require a precoding block length (in time slots) that grows very rapidly with the number of system antennas. [2] Hence, these schemes are not suited for "large" MIMO systems with many BS antennas serving many UTs.

In contrast, Time Division Duplexing (TDD) systems can exploit channel reciprocity for estimating the downlink channels from uplink training. In this case, the system multiplexing gain is still upper bounded by $M^\star(1-M^\star/T)$, but training in the same coherence block is possible (hence, no extra degradation due to prediction) and the training

[1] A channel use corresponds to an independent complex signal-space dimension in the time-frequency domain.

[2] For example, [6] requires precoding over $M!\sum_{j=1}^{M}\frac{1}{j}$ time slots in order to serve M UTs with M BS antennas.

dimension is determined by the number of total UT antennas, while the number of BS antennas can be made as large as desired. By using $M \gg K$ antennas at the BS with TDD, as proposed in [8] (see also the more refined performance analysis and system optimization in [9], [10]), is very attractive for TDD systems both in terms of achieved throughput and in terms of simplified downlink scheduling and signal processing at the BS. Systems where the number of BS antennas are much larger than the number of served UTs are generally referred to as "massive" MIMO. A recent practical testbed implementation of a 64 antenna massive MIMO system, achieving transmitter clock stability and self-calibration in order to effectively exploit TDD reciprocity, has been demonstrated in [11].

In this paper we consider a Joint Spatial Division and Multiplexing (JSDM) approach to potentially achieve massive MIMO-like throughput gains and simplified system operations also for FDD systems, which still represent the far majority of currently deployed cellular networks. We observe that, for a typical cellular configuration, the channel from the M BS antennas to any UT antenna is a *correlated* random vector with covariance matrix that depends on the scattering geometry. Assuming a macro-cellular tower-mounted BS with no significant local scattering, the propagation between the BS antennas and any given UT antenna is characterized by the local scattering around the UT, resulting in the well-known one-ring model [12]. The main idea of JSDM consists of partitioning the user population into groups with *approximately* the same channel covariance eigenspace, and split the downlink beamforming into two stages: a pre-beamforming matrix that depends only on the channel covariances, and a MU-MIMO precoding matrix for the "effective" channel, inclusive of pre-beamforming. The pre-beamforming matrix is chosen in order to minimize the inter-group interference *for any instantaneous channel realization*, by exploiting the linear independence of the *dominant eigenmodes* of the channel covariance matrices of the different groups. Pre-beamforming can be considered as a generalization of *sectorization*, widely used in current cellular technology.

The MU-MIMO precoding stage requires estimation and feedback of the instantaneous (effective) channel realization. As we shall see, this may have significantly reduced dimension with respect to the original physical channel. Therefore, both downlink training and uplink feedback overhead is greatly reduced, making this scheme attractive for FDD systems. Notice that the pre-beamforming stage requires only the channel covariance information, which can be tracked with small protocol overhead.[3]

We show that, under some conditions on the eigenvectors of the channel covariance matrices, JSDM incurs no loss of optimality with respect to the full CSIT case. When these conditions cannot be met, we examine the design of the pre-beamforming matrix and the performance of regularized zero forcing (linear) MU-MIMO precoding for the resulting effective channel. Then, we specialize our system design in the case of Uniform Linear Arrays (ULAs)

[3]In practice, the channel covariance changes over time at a much slower time scale with respect to the system slot rate, therefore we assume that this is locally stationary and can be estimated and tracked using some standard subspace tracking technique [13], [14], [15], [16]. See also the remark at the end of Section III.

and use Szego's asymptotic results on Toeplitz matrices [17] to show that the optimality conditions can be met by ULAs when M is large, as long as the user groups have non-overlapping supports of their Angle of Arrival (AoA) distributions. Using the Toeplitz eigen-subspace approximation result of [17], we argue that the pre-beamforming matrix for large ULAs can be obtained by selecting blocks of columns of a unitary *Discrete Fourier Transform* (DFT) matrix. DFT pre-beamforming achieves very good performance and effective channel dimensionality reduction and requires only a coarse knowledge of the support of the AoA distribution for each user group, without requiring an accurate estimation of the actual channel covariance matrix. Interestingly, related eigen-structure properties of the covariance matrices were independently derived in [18] for the purpose of eliminating the *pilot contamination* effect which limits the performance of TDD massive MIMO with the maximal-ratio single-user beamforming advocated in [8]. Finally, we extend our approach to the case of 2-dimensional ULAs (rectangular antenna arrays) and three-dimensional (3D) beamforming, where we create fixed beams also in the elevation angle direction, in addition to the azimuth angle (planar) direction. The resulting beamforming matrix takes on the appealing form of a Kronecker product. In this way, we can serve simultaneously angular-separated groups of users in different annular regions in a sector, at different distances from the BS. We demonstrate the performance of such a system in a realistic layout assuming a rectangular antenna array mounted on the face of a tall building.

This paper focuses not only on the concept of JSDM, which is not entirely new, but specifically on its performance analysis and system design guidelines, i.e., how to choose the parameters of JSDM for a given set of user groups that we wish to serve simultaneously, on the same time-frequency slot. Since we focus on the large system regime, we can leverage asymptotic random matrix theory results and in particular a recently developed analytical tool referred to as "deterministic equivalent approximation" (see [19] and references therein), which is able to handle the rather complicated class of structured random matrices arising in the JSDM context. Thanks to this analytical tool, all numerical results presented here are obtained in a semi-analytic way, by solving iteratively a provably convergent system of fixed-point equations, without the need of heavy Monte Carlo simulation. For completeness, we provide the equations for the analysis of the basic JSDM schemes without including channel estimation errors in Section V, and in Appendix A the corresponding general case including downlink estimation and noisy CSIT.

Notation : We use boldface capital letters ($\boldsymbol{X}$) for matrices, boldface small letters for vectors ($\boldsymbol{x}$), and small letters (x) for scalars. $\boldsymbol{X}^{\mathsf{T}}$ and $\boldsymbol{X}^{\mathsf{H}}$ denote the transpose and the Hermitian transpose of $\boldsymbol{X}$, $||\boldsymbol{x}||$ denotes the vector 2-norm of $\boldsymbol{x}$, $\mathrm{tr}(\boldsymbol{X})$ and $|\boldsymbol{X}|$ denote the trace and the determinant of the square matrix $\boldsymbol{X}$. The identity matrix is denoted by $\boldsymbol{I}$ (when the dimension is clear from the context) or by $\boldsymbol{I}_n$ (when pointing out its dimension $n \times n$ improves clarity of exposition). $\boldsymbol{X} \otimes \boldsymbol{Y}$ denotes the Kronecker product of two matrices $\boldsymbol{X}, \boldsymbol{Y}$. $\|\boldsymbol{X}\|_F^2 = \mathrm{tr}(\boldsymbol{X}^{\mathsf{H}}\boldsymbol{X})$ indicates the squared Frobenius norm of a matrix $\boldsymbol{X}$. We also use $\mathrm{Span}(\boldsymbol{X})$ to denote the linear subspace generated by columns of $\boldsymbol{X}$ and $\mathrm{Span}^{\perp}(\boldsymbol{X})$ for the orthogonal complement of $\mathrm{Span}(\boldsymbol{X})$. $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that $\boldsymbol{x}$ is a complex

circularly-symmetric Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. Channel Model

We consider the downlink of a single-cell FDD system with a BS with M antennas serving K UTs equipped with a single antenna each. For simplicity, we consider a narrowband (frequency-flat) channel model. By using the Karhunen-Loeve representation, a generic downlink channel vector from the M BS antennas to a UT can be expressed as

$$\boldsymbol{h} = \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{w}, \tag{1}$$

where $\boldsymbol{w} \in \mathbb{C}^{r\times 1} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\Lambda}$ is an $r \times r$ diagonal matrix whose elements are the non-zero eigenvalues of $\boldsymbol{R}$, and $\boldsymbol{U} \in \mathbb{C}^{M\times r}$ is the tall unitary matrix of the eigenvectors of $\boldsymbol{R}$ corresponding to the non-zero eigenvalues. We consider the one-ring model of Fig. 1, where a UT located at azimuth angle θ and distance s is surrounded by a ring of scatterers of radius r such that the AS is $\Delta \approx \arctan(\mathsf{r}/\mathsf{s})$. Assuming a uniform distribution[4] of the received power from planar waves impinging on the BS antennas, the correlation between the channel coefficients of antennas $1 \le m, p \le M$ is given by (see [12] and references therein)

$$[\boldsymbol{R}]_{m,p} = \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} e^{\jmath \boldsymbol{k}^{\mathsf{T}}(\alpha+\theta)(\boldsymbol{u}_m - \boldsymbol{u}_p)} d\alpha, \tag{2}$$

where $\boldsymbol{k}(\alpha) = -\frac{2\pi}{\lambda}(\cos(\alpha), \sin(\alpha))^{\mathsf{T}}$ is the wave vector for a planar wave impinging with AoA α, λ is the carrier wavelength, and $\boldsymbol{u}_m, \boldsymbol{u}_p \in \mathbb{R}^2$ are the vectors indicating the position of BS antennas m, p in the two-dimensional coordinate system (see Fig. 1).

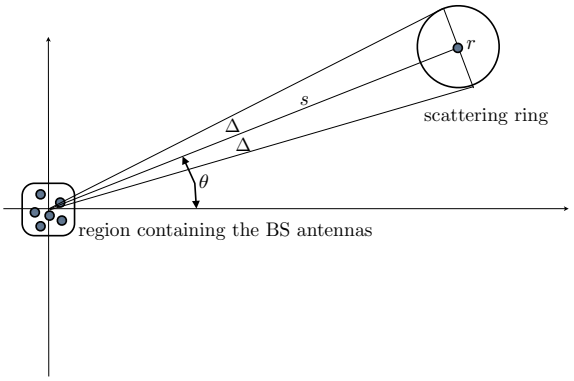


Fig. 1. A UT at AoA θ with a scattering ring of radius r generating a two-sided AS Δ with respect to the BS at the origin.

[4]The uniform distribution is assumed here only for analytical convenience. It is easy to show that similar performances and asymptotic behaviors are achieved by any AoA distribution (measurable non-negative function integrating to 1) with limited support in $[\theta - \Delta, \theta + \Delta]$.

Let $\underline{\boldsymbol{H}}$ denote the $M \times K$ system channel matrix given by stacking the K users channel vectors by columns. The signal vector received by the UTs is given by

$$\boldsymbol{y} = \underline{\boldsymbol{H}}^{\mathsf{H}} \boldsymbol{V} \boldsymbol{d} + \boldsymbol{z} = \underline{\boldsymbol{H}}^{\mathsf{H}} \boldsymbol{x} + \boldsymbol{z} \tag{3}$$

where $\boldsymbol{V}$ is the $M \times S$ precoding matrix with S the rank of the input covariance $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{V}\boldsymbol{d}\boldsymbol{d}^{\mathsf{H}}\boldsymbol{V}^{\mathsf{H}}]$ (i,e., the number of independent data streams sent to the users), $\boldsymbol{d}$ is the S-dimensional transmitted data symbol vector, and $\boldsymbol{z} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ denotes the Gaussian noise at the UT receivers. The transmit signal vector is given by $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{d}$.

III. Joint Spatial Division and Multiplexing

JSDM exploits the fact that, after appropriate partitioning of the UTs such that users in the same group are nearly co-located and different groups are sufficiently well separated in the AoA domain, the structure of the channel covariance matrices can be leveraged in order to reduce the dimensionality of the effective channels and therefore achieve large multiplexing gains with reduced dimension channel training and CSIT feedback.

Suppose that K UTs are selected to form G groups based on the similarity of their channel covariance matrices. We let K_g denote the number of UTs in group g, such that $K = \sum_{g=1}^{G} K_g$, and define the index $g_k = \sum_{g'=1}^{g-1} K_{g'} + k$, for $k = 1, \ldots, K_g$, to denote UT k in group g. Similarly, we let S_g denote the number of independent data streams sent to users in group g, such that $S = \sum_{g=1}^{G} S_g$. We assume for simplicity that all UTs in the same group g have identical covariance matrix $\boldsymbol{R}_g = \boldsymbol{U}_g \Lambda_g \boldsymbol{U}_g^{\mathsf{H}}$, with rank r_g and $r_g^\star \leq r_g$ dominant eigenvalues. In practice, this condition is not verified exactly, but we can select groups such that this condition is closely approximated. Also, the notion of "dominant eigenvalues" is intentionally left fuzzy, since $r_g^\star$ is a design parameter that depends on how much signal power outside the subspace spanned by the corresponding eigenvectors can be tolerated. For future reference, we denote by $\boldsymbol{U}_g^\star$ the $M \times r_g^\star$ matrix collecting the dominant eigenvectors, and let $\boldsymbol{U}_g = [\boldsymbol{U}_g^\star, \boldsymbol{U}_g']$, with $\boldsymbol{U}_g'$ of dimension $M \times (r_g - r_g^\star)$, containing the eigenvectors corresponding to the weakest eigenvalues. Notice that, by construction, we have that $0 \leq S_g \leq \min\{K_g, r_g^\star\}$, since we cannot deliver more independent symbol streams than the multiplexing gain $\min\{K_g, r_g^\star\}$ of each group g.

The channel vector of user g_k is given by $\boldsymbol{h}_{g_k} = \boldsymbol{U}_g \boldsymbol{\Lambda}_g^{\frac{1}{2}} \boldsymbol{w}_{g_k}$. We let $\boldsymbol{H}_g = \left[\boldsymbol{h}_{g_1}, \cdots, \boldsymbol{h}_{g_{K_g}}\right]$ and $\underline{\boldsymbol{H}} = \left[\boldsymbol{H}_1, \cdots, \boldsymbol{H}_G\right]$ denote the group g channel matrix and the overall system channel matrix, respectively. As anticipated in Section I, JSDM is based on two-stage precoding. Namely, we let $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{P}$, where $\boldsymbol{B} \in \mathbb{C}^{M \times b}$ is a *pre-beamforming* matrix, $\boldsymbol{P} \in \mathbb{C}^{b \times S}$ is a MU-MIMO precoding matrix, and where $b \geq S$ is an integer design parameter, to be optimized. The pre-beamforming matrix $\boldsymbol{B}$ is a function of the channels second-order statistics, i.e., it depends on the set $\{\boldsymbol{U}_g, \boldsymbol{\Lambda}_g\}$, or on some directional information extracted from the channel covariance matrices (AoA and AS of the different groups). In any case, $\boldsymbol{B}$ is independent of the instantaneous realization of the channel matrix $\underline{\boldsymbol{H}}$. The MU-MIMO precoding matrix $\boldsymbol{P}$ is allowed to depend on the instantaneous realization of the reduced dimensional *effective channel*

$\underline{\mathbf{H}} \triangleq \boldsymbol{B}^{\mathsf{H}}\underline{\boldsymbol{H}}$. We let $b = \sum_{g=1}^{G} b_g$ such that $b_g \geq S_g$, and let $\boldsymbol{B}_g$ be the $M \times b_g$ pre-beamforming matrix of group g. The received signal (3) can be rewritten as

$$\boldsymbol{y} = \underline{\mathbf{H}}^{\mathsf{H}}\boldsymbol{P}\boldsymbol{d} + \boldsymbol{z} \tag{4}$$

where

$$\underline{\mathbf{H}}^{\mathsf{H}} = \begin{bmatrix} \boldsymbol{H}_1^{\mathsf{H}}\boldsymbol{B}_1 & \boldsymbol{H}_1^{\mathsf{H}}\boldsymbol{B}_2 & \cdots & \boldsymbol{H}_1^{\mathsf{H}}\boldsymbol{B}_G \\ \boldsymbol{H}_2^{\mathsf{H}}\boldsymbol{B}_1 & \boldsymbol{H}_2^{\mathsf{H}}\boldsymbol{B}_2 & \cdots & \boldsymbol{H}_2^{\mathsf{H}}\boldsymbol{B}_G \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_G^{\mathsf{H}}\boldsymbol{B}_1 & \boldsymbol{H}_G^{\mathsf{H}}\boldsymbol{B}_2 & \cdots & \boldsymbol{H}_G^{\mathsf{H}}\boldsymbol{B}_G \end{bmatrix},$$

and where $\boldsymbol{H}_g^{\mathsf{H}}\boldsymbol{B}_{g'}$ is the $K_g \times b_{g'}$ effective channel matrix connecting the users of group g with the effective channel inputs of group g'.

If the estimation and feedback of the effective channel $\underline{\mathbf{H}}$ can be afforded, the precoding matrix $\boldsymbol{P}$ is determined as a function of the whole $\underline{\mathbf{H}}$. We refer to this approach as *Joint Group Processing* (JGP). However, this may still be too costly in terms of transmission resource. Hence, a lower complexity and generally more attractive approach consists of estimating and feeding back only the G diagonal blocks $\mathbf{H}_g = \boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{H}_g$, of dimension $b_g \times K_g$, and treating each group separately. We refer to this approach as *Per-Group Processing* (PGP). In this case, the precoding matrix takes on the block-diagonal form $\boldsymbol{P} = \mathrm{diag}(\boldsymbol{P}_1, \cdots, \boldsymbol{P}_G)$, where $\boldsymbol{P}_g \in \mathbb{C}^{b_g \times S_g}$, resulting in the vector broadcast plus interference Gaussian channel

$$\boldsymbol{y}_g = \mathbf{H}_g^{\mathsf{H}}\boldsymbol{P}_g\boldsymbol{d}_g + \sum_{g' \neq g} \boldsymbol{H}_g{}^{\mathsf{H}}\boldsymbol{B}_{g'}\boldsymbol{P}_{g'}\boldsymbol{d}_{g'} + \boldsymbol{z}_g, \quad \text{for } g = 1, \ldots, G. \tag{5}$$

With PGP, it is interesting to choose the groups and design the pre-beamforming matrix such that, with high probability,

$$\boldsymbol{H}_g{}^{\mathsf{H}}\boldsymbol{B}_{g'} \approx \mathbf{0}, \text{ for all } g' \neq g. \tag{6}$$

Exact Block Diagonalization (BD) is possible if $\mathrm{Span}(\boldsymbol{U}_g) \nsubseteq \mathrm{Span}(\{\boldsymbol{U}_{g'} : g' \neq g\})$ for all $g = 1, \ldots, G$. In particular, multiplexing gain S_g (i.e., the number of interference-free data streams) can be achieved for group g if and only if

$$\dim\left(\mathrm{Span}(\boldsymbol{U}_g) \cap \mathrm{Span}^{\perp}(\{\boldsymbol{U}_{g'} : g' \neq g\})\right) \geq S_g. \tag{7}$$

Approximate BD can be achieved by selecting $r_g^{\star}$ dominant eigenmodes for each group g, such that $\mathrm{Span}(\boldsymbol{U}_g^{\star}) \nsubseteq \mathrm{Span}(\{\boldsymbol{U}_{g'}^{\star} : g' \neq g\})$ for all $g = 1, \ldots, G$. In this case, in order to deliver S_g streams to group g we require

$$\dim\left(\mathrm{Span}(\boldsymbol{U}_g^{\star}) \cap \mathrm{Span}^{\perp}(\{\boldsymbol{U}_{g'}^{\star} : g' \neq g\})\right) \geq S_g. \tag{8}$$

However, these streams will be affected by some residual interference due to the weak eigenmodes not included in the matrices $\{\boldsymbol{U}_g^{\star} : g = 1, \ldots, G\}$.

Remark 1: Notice that the PGP pre-beamforming creates *virtual sectors*, i.e., a generalization of spatial sectorization commonly used in current cellular technology. Each group corresponds to a virtual sector, and it is independently precoded under a total sum power constraint, possibly incurring some residual *inter-group interference* in the case of approximate BD. ♢

Remark 2: It is reasonable to assume that the channel covariance matrix $\boldsymbol{R}_g$ for each user group changes slowly with respect to the coherence time of the instantaneous channel matrix $\boldsymbol{H}_g$. The dominant eigenmodes $\boldsymbol{U}_g^\star$ can be tracked for each UT using a suitable subspace estimation and tracking algorithm [20], by exploiting the downlink training phase, and they can be fed back to the BS at a low rate. Furthermore, for particularly designed BS antenna configurations, these estimates can be refined at the BS by exploiting the uplink, even though in an FDD system this takes place at a different carrier frequency (see for example [21]). The estimation and tracking of the (slowly time-varying) channel statistics is a topic of great interest in this context, but it is out of the scope of this paper. Here, we assume that the channel covariance matrix for each user is known. ♢

IV. JSDM WITH EIGEN-BEAMFORMING

A. Achieving capacity with reduced CSIT

Let $r = \sum_{g=1}^{G} r_g$ and suppose that the channel covariances of the G groups are such that $\underline{\boldsymbol{U}} = [\boldsymbol{U}_1, \cdots, \boldsymbol{U}_G]$ is $M \times r$ *tall unitary* (i.e., $r \leq M$ and $\underline{\boldsymbol{U}}^{\mathsf{H}}\underline{\boldsymbol{U}} = \boldsymbol{I}_r$). In order to obtain *exact* BD it is sufficient to let $b_g = r_g$ and $\boldsymbol{B}_g = \boldsymbol{U}_g$. This choice for the pre-beamforming matrix is referred to in the following as *eigen-beamforming*. In this case, the decoupled MU-MIMO channel (5) takes on the form

$$\boldsymbol{y}_g = \mathsf{H}_g{}^{\mathsf{H}}\boldsymbol{P}_g\boldsymbol{d}_g + \boldsymbol{z}_g = \boldsymbol{W}_g^{\mathsf{H}}\Lambda_g^{1/2}\boldsymbol{P}_g\boldsymbol{d}_g + \boldsymbol{z}_g, \qquad \text{for } \; g = 1, \ldots, G, \tag{9}$$

where $\boldsymbol{W}_g$ is a $r_g \times K_g$ i.i.d. matrix with elements $\sim \mathcal{CN}(0,1)$. In this case we have:

Theorem 1: For $\underline{\boldsymbol{U}}$ tall unitary, JSDM with PGP achieves the same sum capacity of the corresponding MU-MIMO downlink channel (3) with full CSIT.

Proof: Let $\mathcal{C}^{\mathrm{sum}}(\underline{\boldsymbol{H}}; P)$ denote the sum capacity of (3) with sum power constraint P and fixed channel matrix $\underline{\boldsymbol{H}}$, perfectly known to transmitter and receivers. By the MAC-BC duality [22], we have

$$\mathcal{C}^{\mathrm{sum}}(\underline{\boldsymbol{H}}; P) = \max_{\boldsymbol{S}_g \succeq 0 : \sum_g \mathrm{tr}(\boldsymbol{S}_g) \leq P} \; \log \left| \boldsymbol{I}_M + \sum_{g=1}^{G} \boldsymbol{U}_g \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{W}_g \boldsymbol{S}_g \boldsymbol{W}_g^{\mathsf{H}} \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{U}_g^{\mathsf{H}} \right| \tag{10}$$

where $\boldsymbol{S}_g$ denotes the diagonal $K_g \times K_g$ input covariance matrix for group g in the dual MAC channel. For any fixed set $\{\boldsymbol{S}_g\}$ of feasible input covariance matrices, define for notation simplicity $\boldsymbol{A}_g = \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{W}_g \boldsymbol{S}_g \boldsymbol{W}_g^{\mathsf{H}} \boldsymbol{\Lambda}_g^{1/2}$. Notice that

$\boldsymbol{A}_g$ has dimension $r_g \times r_g$ and is invertible with probability 1 over the random channel realization. The theorem is proved by showing the the determinant identity

$$\left| \boldsymbol{I}_M + \sum_{g=1}^{G} \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right| = \prod_{g=1}^{G} \left| \boldsymbol{I}_M + \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right|. \tag{11}$$

This can be proved by induction, noticing the following step: for any $1 \leq g' \leq G$,

$$\begin{aligned} \left| \boldsymbol{I}_M + \sum_{g=g'}^{G} \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right| &= \left| \boldsymbol{I}_M + \boldsymbol{U}_{g'} \boldsymbol{A}_{g'} \boldsymbol{U}_{g'}^{\mathsf{H}} \right| \left| \boldsymbol{I}_M + (\boldsymbol{I}_M + \boldsymbol{U}_{g'} \boldsymbol{A}_{g'} \boldsymbol{U}_{g'}^{\mathsf{H}})^{-1} \sum_{g=g'+1}^{G} \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right| \\ &= \left| \boldsymbol{I}_M + \boldsymbol{U}_{g'} \boldsymbol{A}_{g'} \boldsymbol{U}_{g'}^{\mathsf{H}} \right| \left| \boldsymbol{I}_M + (\boldsymbol{I}_M - \boldsymbol{U}_{g'} (\boldsymbol{A}_{g'}^{-1} + \boldsymbol{I}_{r_g})^{-1} \boldsymbol{U}_{g'}^{\mathsf{H}}) \sum_{g=g'+1}^{G} \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right| \qquad (12) \\ &= \left| \boldsymbol{I}_M + \boldsymbol{U}_{g'} \boldsymbol{A}_{g'} \boldsymbol{U}_{g'}^{\mathsf{H}} \right| \left| \boldsymbol{I}_M + \sum_{g=g'+1}^{G} \boldsymbol{U}_g \boldsymbol{A}_g \boldsymbol{U}_g^{\mathsf{H}} \right|, \qquad (13) \end{aligned}$$

where (12) follows form the matrix inversion lemma and (13) follows from the the fact that, by assumption, $\boldsymbol{U}_{g'}^{\mathsf{H}} \boldsymbol{U}_g = \boldsymbol{0}$ for all $g' \neq g$. Using (11) in (10) we obtain

$$\mathcal{C}^{\mathrm{sum}}(\underline{\boldsymbol{H}}; P) = \max_{\boldsymbol{S}_g \succeq 0 : \sum_g \mathrm{tr}(\boldsymbol{S}_g) \leq P} \sum_{g=1}^{G} \log \left| \boldsymbol{I}_{r_g} + \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{W}_g \boldsymbol{S}_g \boldsymbol{W}_g^{\mathsf{H}} \boldsymbol{\Lambda}_g^{1/2} \right|, \tag{14}$$

which is immediately recognized to be the capacity of the dual MAC (with sum power constraint) for the set of decoupled MU-MIMO downlink channels (9). ■

Remark 3: In a similar manner it is possible to show that under the orthogonality condition of Theorem 1, JSDM achieves the whole capacity region [23], and not only the sum capacity. In order to see this, for any user subset $\mathcal{K} \subseteq \{1, \ldots, K\}$ define $\boldsymbol{H}_g(\mathcal{K})$ as the sub matrix of $\boldsymbol{H}_g$ obtained by selecting the columns $g_k \in \mathcal{K}$, and let $\boldsymbol{S}_g(\mathcal{K})$ denote the submatrix of $\boldsymbol{S}_g$ obtained by retaining the rows and columns corresponding to users $g_k \in \mathcal{K}$. Then, the capacity region of the dual MAC of (3) subject to the sum power constraint can be written as

$$\mathcal{C}(\underline{\boldsymbol{H}}; P) = \bigcup_{\substack{\boldsymbol{S}_g \succeq 0 : \\ \sum_{g=1}^{G} \mathrm{Tr}(\boldsymbol{S}_g) \leq P}} \left\{ \boldsymbol{r} \in \mathbb{R}_K^+ : \sum_{g_k \in \mathcal{K}} r_{g_k} \leq \log \left| \boldsymbol{I}_M + \sum_{g=1}^{G} \boldsymbol{H}_g(\mathcal{K}) \boldsymbol{S}_g(\mathcal{K}) \boldsymbol{H}_g^{\mathsf{H}}(\mathcal{K}) \right|, \ \forall\, \mathcal{K} \subseteq \{1, \ldots, K\} \right\}. \tag{15}$$

The determinant identity (11) can be applied to the partial sum-rate bounds for each user subset $\mathcal{K}$, noticing that the tall unitary condition of the singular vectors is retained by the new system matrix $\underline{\boldsymbol{H}}(\mathcal{K}) = [\boldsymbol{H}_1(\mathcal{K}), \ldots, \boldsymbol{H}_G(\mathcal{K})]$. ◊

Remark 4: Theorem 1 has an important practical implication: in a situation where a large number of UTs, each of which has its own AoA and AS, must be served by the downlink, a good scheduling strategy consists of the following. First, partition the users into groups with (approximately) identical eigenspaces. Then, partition the collection of groups into disjoint and mutually exclusive sets, such that the groups in each set satisfy the tall unitary condition of Theorem 1, and such that the number of sets is minimal, over all possible partitions. Finally, schedule the groups in each set

to be served simultaneously, on the same time-frequency slot, using JSDM, and use time-frequency sharing across the groups. Notice that this does not mean that, in general, JSDM is optimal. In fact, in order to meet the tall unitary condition we may be obliged to reduce the number G of simultaneously served groups in each set. As already noticed for the problem of clustering users into groups, also the problem of finding optimal partitions of the user groups under JSDM with PGP is far from trivial, and goes beyond the scope of this paper. $\lozenge$

When achieving the tall unitary condition is too restrictive in terms of multiplexing gain, the pre-beamforming matrix $\boldsymbol{B}$ can be chosen as a function of the whole $\underline{\boldsymbol{U}}$ in order to achieve exact or approximated BD. This approach is presented in the next section.

B. Block diagonalization

Recall that $\boldsymbol{B} = [\boldsymbol{B}_1, \ldots, \boldsymbol{B}_G]$ is an $M \times b$ matrix consisting of G blocks of dimension $M \times b_g$, each corresponding to a particular group g. For given target numbers of streams per group $\{S_g\}$ and dimensions $\{b_g\}$ satisfying $S_g \leq b_g \leq r_g$, our goal is to design the blocks $\boldsymbol{B}_g$ such that BD is achieved, i.e., $\boldsymbol{U}_{g'}^{\mathsf{H}} \boldsymbol{B}_g = \boldsymbol{0}$ for all $g' \neq g$ and $\mathrm{rank}(\boldsymbol{U}_g^{\mathsf{H}} \boldsymbol{B}_g) \geq S_g$. A necessary condition for exact zero-forcing of the off-diagonal blocks is $\mathrm{Span}(\boldsymbol{B}_g) \subseteq \mathrm{Span}^{\perp}(\{\boldsymbol{U}_{g'} : g' \neq g\})$. When $\mathrm{Span}^{\perp}(\{\boldsymbol{U}_{g'} : g' \neq g\})$ has dimension smaller than S_g, the rank condition on the diagonal blocks cannot be satisfied. In this case, S_g should be reduced or, as an alternative, approximated BD based on selecting $r_g^{\star} < r_g$ dominant eigenmodes for each group g can be implemented. This consists of replacing $\boldsymbol{U}_g$ with $\boldsymbol{U}_g^{\star}$ in the above conditions. When $\mathrm{Span}(\{\boldsymbol{U}_{g'} : g' \neq g\})$ has dimension M, then exact BD cannot be achieved even for $S_g = 1$, and therefore approximated BD should be considered in any case. Without loss of generality, we formulate the design of $\{\boldsymbol{B}_g\}$ for approximated BD with some feasible choice of the parameters $\{r_g^{\star}\}$, $\{b_g\}$ and $\{S_g\}$. It should be noticed that these are design parameters that should be optimized for a given system configuration, in order to maximize the overall spectral efficiency. This optimization is far from trivial. For the time being, we consider an arbitrary feasible choice and postpone the discussion on the tradeoff that governs the design of these parameters in Sections V-C (see Remark 5) and VI-A (see Remark 7).

Following the approach of [24], we define

$$\boldsymbol{\Xi}_g = \left[\boldsymbol{U}_1^{\star}, \ldots, \boldsymbol{U}_{g-1}^{\star}, \boldsymbol{U}_{g+1}^{\star}, \ldots, \boldsymbol{U}_G^{\star}\right], \tag{16}$$

of dimensions $M \times \sum_{g' \neq g} r_{g'}^{\star}$ and rank $\sum_{g' \neq g} r_{g'}^{\star}$, and let $[\boldsymbol{E}_g^{(1)}, \boldsymbol{E}_g^{(0)}]$ denote a system of left eigenvectors of $\boldsymbol{\Xi}_g$ (e.g., obtained by Singular Value Decomposition (SVD)), such that $\boldsymbol{E}_g^{(0)}$ is $M \times \left(M - \sum_{g' \neq g} r_{g'}^{\star}\right)$ and forms a unitary basis for the orthogonal complement of $\mathrm{Span}(\boldsymbol{\Xi}_g)$, i.e., such that $\mathrm{Span}(\boldsymbol{E}_g^{(0)}) = \mathrm{Span}^{\perp}(\{\boldsymbol{U}_{g'}^{\star} : g' \neq g\})$.

We obtain $\boldsymbol{B}_g$ by concatenating the projection onto $\mathrm{Span}(\boldsymbol{E}_g^{(0)})$ with eigen-beamforming along the dominant eigenmodes of the covariance matrix of the resulting projected channels of group g, i.e., of the columns of $(\boldsymbol{E}_g^{(0)})^{\mathsf{H}} \boldsymbol{H}_g$. Recalling the Karhunen-Loeve decomposition (1), we have that the covariance matrix of $\widehat{\boldsymbol{h}}_{g_k} = (\boldsymbol{E}_g^{(0)})^{\mathsf{H}} \boldsymbol{U}_g \boldsymbol{\Lambda}_g^{1/2} \boldsymbol{w}_{g_k}$

is given by

$$\widehat{\boldsymbol{R}}_g = (\boldsymbol{E}_g^{(0)})^{\mathsf{H}} \boldsymbol{U}_g \boldsymbol{\Lambda}_g \boldsymbol{U}_g^{\mathsf{H}} \boldsymbol{E}_g^{(0)} = \boldsymbol{G}_g \boldsymbol{\Phi}_g \boldsymbol{G}_g^{\mathsf{H}}, \tag{17}$$

where the expression on the right of (17) is the SVD of $\widehat{\boldsymbol{R}}_g$. Letting $\boldsymbol{G}_g = [\boldsymbol{G}_g^{(1)}, \boldsymbol{G}_g^{(0)}]$ where $\boldsymbol{G}_g^{(1)}$ contains the dominant b_g eigenmodes of $\widehat{\boldsymbol{R}}_g$, we eventually obtain

$$\boldsymbol{B}_g = \boldsymbol{E}_g^{(0)} \boldsymbol{G}_g^{(1)}. \tag{18}$$

The pre-beamforming matrix $\boldsymbol{B}_g$ can be interpreted as being orthogonal to the dominant $r_{g'}^*$ eigenmodes of groups $g' \neq g$, and matched to the b_g dominant eigenmodes of the covariance matrix of the projected channels $(\boldsymbol{E}_g^{(0)})^{\mathsf{H}} \boldsymbol{H}_g$ of group g. By construction, we have that b_g is less or equal to the rank of $\widehat{\boldsymbol{R}}_g$, given by $\min\left\{r_g, M - \sum_{g' \neq g} r_{g'}^\star\right\}$. In particular, if $r = \sum_g r_g \leq M$, we can choose $b_g = r_g^\star = r_g$ and obtain exact BD.

V. Performance analysis with linear precoding

In this section we provide expressions for the performance analysis of JSDM with JGP and PGP and linear precoding, using the techniques of deterministic equivalents [10]. For simplicity of exposition, we consider a symmetric scenario with the same number $K_g = K'$ of users per group, the same number $S_g = S'$ of streams per group, and the same dimension $b_g = b'$ of the pre-beamforming matrix per group. However, the analysis can be immediately extended to the general case considered before. This technique can be applied as long as the users to be served in each group are selected independently of their instantaneous channel realization. Hence, we assume that for each group a subset of S' out of the possible K' users is pre-selected and scheduled for transmission over the current downlink time-frequency slot. This simplified scheduling requires only the instantaneous CSIT feedback from the pre-scheduled users [5] and it is in line with the massive MIMO concept, where hardware augmentation at the BS allows significant simplification in the system operations.

Under these assumptions, the transformed channel matrix $\underline{\mathsf{H}}$ has dimension $b \times S$, with blocks H_g of dimension $b' \times S'$. Also for the sake of simplicity and in line with massive MIMO system simplification (see for example [8], [9]) we allocate to all users the same fraction of the total transmit power P, such that the data vector covariance matrix is given by $\mathbb{E}[\boldsymbol{d}\boldsymbol{d}^{\mathsf{H}}] = \frac{P}{S}\boldsymbol{I}_S$. In the following, we present the deterministic equivalent fixed-point equations for determining the *Signal-to-Interference plus Noise Ratio* (SINR) at the UTs receivers for the case of JSDM with JGP and PGP with linear *regularized zero forcing precoding*. Along the same lines, Appendix A presents the case of regularized and non-regularized linear zero forcing precoding for PGP in the case of noisy CSIT obtained from downlink training (see Section VI). It is well-known that a discrete-time complex additive noise plus interference

[5]Unlike channel-based opportunistic user selection, [25]–[28], that requires to collect CSIT from many users and then select a subset of users with quasi-orthogonal channel vectors.

channel with SINR equal to γ has capacity at least as large as $\log(1+\gamma)$ bit/symbol [29]. Hence, in order to obtain an asymptotically convergent approximation of the achievable spectral efficiency (in bit/symbol) per served user, we compute γ via the deterministic equivalent method, and plug the result into the $\log(1+\gamma)$ rate formula.

A. JSDM with joint group processing

For fixed pre-beamforming matrix $\boldsymbol{B}$ and JGP, the regularized zero forcing precoding matrix is given by

$$\boldsymbol{P}_{\text{rzf}} = \zeta \boldsymbol{K} \underline{\mathbf{H}}, \tag{19}$$

where $\boldsymbol{K} = \left[\underline{\mathbf{H}}\underline{\mathbf{H}}^{\mathsf{H}} + b\alpha \boldsymbol{I}_b\right]^{-1}$, α is a regularization factor, and ζ is a normalization factor chosen to satisfy the power constraint and is given by

$$\zeta^2 = \frac{S}{\text{tr}\left(\boldsymbol{P}_{\text{rzf}}^{\mathsf{H}} \boldsymbol{B}^{\mathsf{H}} \boldsymbol{B} \boldsymbol{P}_{\text{rzf}}\right)}. \tag{20}$$

The covariance matrix of the transformed channel of group g is given by

$$\tilde{\boldsymbol{R}}_g = \begin{bmatrix} \boldsymbol{B}_1^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_1 & \boldsymbol{B}_1^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_1^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_G \\ \boldsymbol{B}_2^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_1 & \boldsymbol{B}_2^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_2^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_G \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_G^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_1 & \boldsymbol{B}_G^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_G^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_G \end{bmatrix}. \tag{21}$$

The SINR for user g_k is given by

$$\gamma_{g_k,\text{jgp,rzf}} = \frac{\frac{P}{S}\zeta^2 |\boldsymbol{h}_{g_k}^{\mathsf{H}} \boldsymbol{B} \boldsymbol{K} \boldsymbol{B}^{\mathsf{H}} \boldsymbol{h}_{g_k}|^2}{\frac{P}{S}\sum_{j \neq g_k} \zeta^2 |\boldsymbol{h}_{g_k}^{\mathsf{H}} \boldsymbol{B} \boldsymbol{K} \boldsymbol{B}^{\mathsf{H}} \boldsymbol{h}_j|^2 + 1} \tag{22}$$

where the subscript "jgp" stands for *joint group processing*.

Following the approach of [10], assuming that as $M \to \infty$ the other system dimensions r, S and b also go to infinity linearly with M, we have

$$\gamma_{g_k,\text{jgp,rzf}} - \gamma^o_{g_k,\text{jgp,rzf}} \overset{M\to\infty}{\longrightarrow} 0 \quad \text{with probability } 1, \tag{23}$$

where, for all users g_k, $\gamma^o_{g_k,\text{jgp,rzf}}$ is a deterministic quantity that can be computed for any finite M as

$$\gamma^o_{g_k,\text{jgp,rzf}} = \frac{\frac{P}{S}\zeta^2 (m^o_g)^2}{\zeta^2 \Upsilon^o_g + (1+m^o_g)^2}, \tag{24}$$

where $\zeta^2 = \frac{P}{\Gamma^o}$ and the quantities m_g^o, Υ_g^o and Γ^o are obtained by solving the system of fixed-point equations

$$m_g^o = \frac{1}{b}\text{tr}\left(\tilde{\boldsymbol{R}}_g \boldsymbol{T}\right) \tag{25}$$

$$\boldsymbol{T} = \left(\frac{S'}{b}\sum_{g=1}^{G}\frac{\tilde{\boldsymbol{R}}_g}{1+m_g^o} + \alpha \boldsymbol{I}_b\right)^{-1} \tag{26}$$

$$\Gamma^o = \frac{1}{b}\frac{P}{G}\sum_{g=1}^{G}\frac{n_g}{(1+m_g^o)^2} \tag{27}$$

$$\Upsilon_g^o = \frac{1}{b}\frac{P}{G}\left[\sum_{g'=1,g'\neq g}^{G}\frac{n_{g',g}}{(1+m_{g'}^o)^2} + \frac{S'-1}{S'}\frac{n_{g,g}}{(1+m_g^o)^2}\right], \tag{28}$$

with $\boldsymbol{n} = [n_1, n_2, \ldots, n_G]^{\mathsf{T}}$ and $\boldsymbol{n}_g = [n_{1,g}, n_{2,g}, \ldots, n_{G,g}]^{\mathsf{T}}$ defined by

$$\boldsymbol{n} = (\boldsymbol{I}_G - \boldsymbol{J})^{-1}\boldsymbol{v} \tag{29}$$

$$\boldsymbol{n}_g = (\boldsymbol{I}_G - \boldsymbol{J})^{-1}\boldsymbol{v}_g, \tag{30}$$

where $\boldsymbol{J}, \boldsymbol{v}$ and $\boldsymbol{v}_g$ are given as

$$[\boldsymbol{J}]_{g,g'} = \frac{\frac{S'}{b}\text{tr}\left(\tilde{\boldsymbol{R}}_g \boldsymbol{T}\tilde{\boldsymbol{R}}_{g'}\boldsymbol{T}\right)}{b(1+m_{g'}^o)^2} \tag{31}$$

$$\boldsymbol{v} = \frac{1}{b}\left[\text{tr}\left(\tilde{\boldsymbol{R}}_1 \boldsymbol{T}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B}\boldsymbol{T}\right), \ldots, \text{tr}\left(\tilde{\boldsymbol{R}}_G \boldsymbol{T}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B}\boldsymbol{T}\right)\right]^{\mathsf{T}} \tag{32}$$

$$\boldsymbol{v}_g = \frac{1}{b}\left[\text{tr}\left(\tilde{\boldsymbol{R}}_1 \boldsymbol{T}\tilde{\boldsymbol{R}}_g\boldsymbol{T}\right), \ldots, \text{tr}\left(\tilde{\boldsymbol{R}}_G \boldsymbol{T}\tilde{\boldsymbol{R}}_g\boldsymbol{T}\right)\right]^{\mathsf{T}} \tag{33}$$

B. JSDM with per-group processing

The channel covariance matrix for a user g_k is given by $\bar{\boldsymbol{R}}_g = \boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{R}_g\boldsymbol{B}_g$. Focusing only on the users in group g, the regularized zero forcing precoding matrix is given by

$$\boldsymbol{P}_{g,\text{rzf}} = \bar{\zeta}_g \bar{\boldsymbol{K}}_g \mathbf{H}_g, \tag{34}$$

where $\bar{\boldsymbol{K}}_g = \left[\mathbf{H}_g\mathbf{H}_g^{\mathsf{H}} + b'\alpha\boldsymbol{I}_{b'}\right]^{-1}$, α is a regularization factor, and $\bar{\zeta}_g$ is the power normalization factor given by

$$\bar{\zeta}_g^2 = \frac{S'}{\text{tr}\left(\boldsymbol{P}_{g,\text{rzf}}^{\mathsf{H}}\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g\boldsymbol{P}_{g,\text{rzf}}\right)}. \tag{35}$$

When $\boldsymbol{B}_g$ is given by (18), then it is the product of two tall unitary matrices so that $\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g = \boldsymbol{I}_{b'}$. However, we use (35) for the sake of generality.

The SINR of user g_k given by

$$\gamma_{g_k,\text{pgp}} = \frac{\frac{P}{S}\bar{\zeta}_g^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_g\bar{\boldsymbol{K}}_g\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{h}_{g_k}|^2}{\frac{P}{S}\sum_{j\neq k}\bar{\zeta}_g^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_g\bar{\boldsymbol{K}}_g\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{h}_{g_j}|^2 + \frac{P}{S}\sum_{g'\neq g}\sum_j\bar{\zeta}_{g'}^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_{g'}\bar{\boldsymbol{K}}_{g'}\boldsymbol{B}_{g'}^{\mathsf{H}}\boldsymbol{h}_{g'_j}|^2 + 1} \tag{36}$$

where the subscript "pgp" stands for *per-group processing*.

Proceeding similarly as before and applying the method developed in [10], and assuming that as $M \to \infty$ the other system dimensions r, S and b also go to infinity linearly with M, we have

$$\gamma_{g_k,\mathrm{pgp,rzf}} - \gamma^o_{g_k,\mathrm{pgp,rzf}} \overset{M\to\infty}{\longrightarrow} 0 \quad \text{with probability 1}, \tag{37}$$

where, for all users g_k, $\gamma^o_{g_k,\mathrm{pgp,rzf}}$ is a deterministic quantity that can be computed for any finite M as

$$\gamma^o_{g_k,\mathrm{pgp,rzf}} = \frac{\frac{P}{S}\bar{\zeta}^2_g(\bar{m}^o_g)^2}{\bar{\zeta}^2_g\bar{\Upsilon}^o_{g,g} + (1+\sum_{g'\neq g}\bar{\zeta}^2_{g'}\bar{\Upsilon}^o_{g,g'})(1+\bar{m}^o_g)^2}, \tag{38}$$

where $\bar{\zeta}^2_g = \frac{P/G}{\bar{\Gamma}^o_g}$ and the quantities $\bar{m}^o_g$, $\bar{\Upsilon}^o_{g,g}$, $\bar{\Upsilon}^o_{g,g'}$ and $\bar{\Gamma}^o_g$ are given by

$$\bar{m}^o_g = \frac{1}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\right) \tag{39}$$

$$\bar{\boldsymbol{T}}_g = \left(\frac{S'}{b'}\frac{\bar{\boldsymbol{R}}_g}{1+\bar{m}^o_g} + \alpha\boldsymbol{I}_{b'}\right)^{-1} \tag{40}$$

$$\bar{\Gamma}^o_g = \frac{1}{b'}\frac{P}{G}\frac{\bar{n}_g}{(1+\bar{m}^o_g)^2} \tag{41}$$

$$\bar{\Upsilon}^o_{g,g} = \frac{1}{b'}\frac{S'-1}{S'}\frac{P}{G}\frac{\bar{n}_{g,g}}{(1+\bar{m}^o_g)^2} \tag{42}$$

$$\bar{\Upsilon}^o_{g,g'} = \frac{1}{b'}\frac{P}{G}\frac{\bar{n}_{g',g}}{(1+\bar{m}^o_{g'})^2} \tag{43}$$

$$\bar{n}_g = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\boldsymbol{B}^{\mathsf{H}}_g\boldsymbol{B}_g\bar{\boldsymbol{T}}_g\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\right)}{b'(1+\bar{m}^o_g)^2}} \tag{44}$$

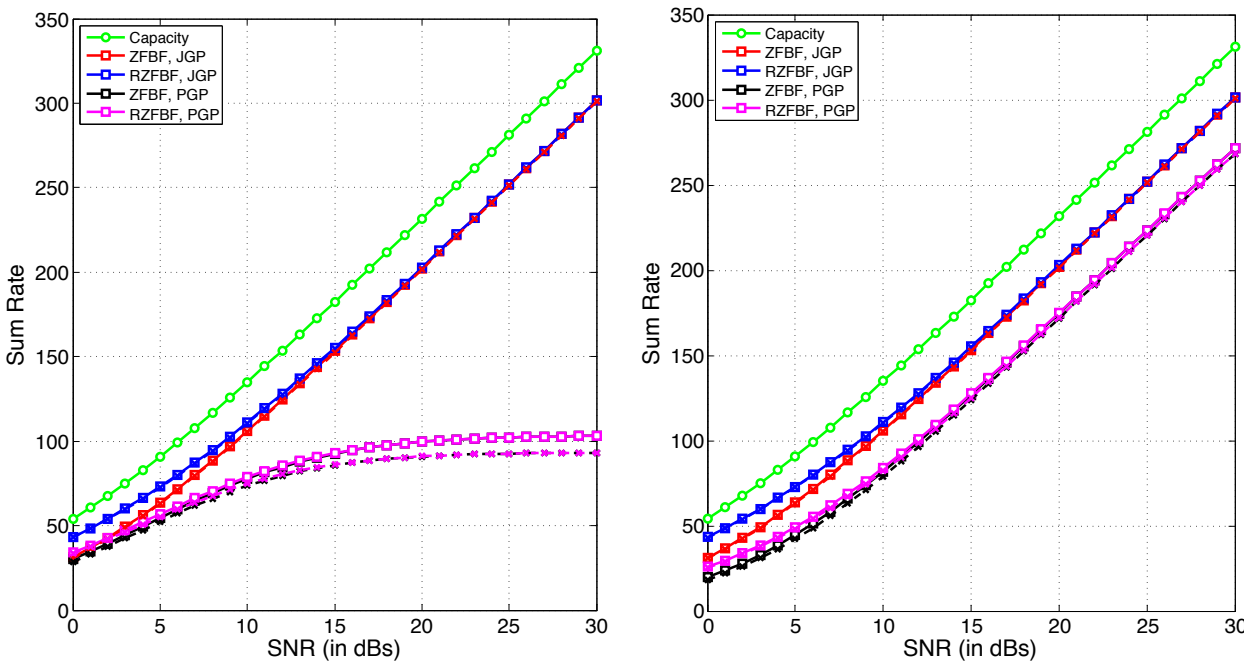
$$\bar{n}_{g,g} = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\bar{\boldsymbol{R}}_g\bar{\boldsymbol{T}}_g\right)}{b'(1+\bar{m}^o_g)^2}} \tag{45}$$

$$\bar{n}_{g',g} = \frac{\frac{1}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_{g'}\bar{\boldsymbol{T}}_{g'}\boldsymbol{B}^{\mathsf{H}}_{g'}\boldsymbol{R}_g\boldsymbol{B}_{g'}\bar{\boldsymbol{T}}_{g'}\right)}{1-\frac{\frac{S'}{b'}\mathrm{tr}\left(\bar{\boldsymbol{R}}_{g'}\bar{\boldsymbol{T}}_{g'}\bar{\boldsymbol{R}}_{g'}\bar{\boldsymbol{T}}_{g'}\right)}{b'(1+\bar{m}^o_{g'})^2}} \tag{46}$$

C. Validation of the asymptotic analysis

In this section we present some numerical examples focusing on the case when the tall unitary condition is not satisfied, and we discuss the choice of the effective rank parameter $r^\star$ in the approximated BD for PGP (more in general, the parameters $\{r^\star_g\}$, for an asymmetric case). We also compare the results obtained via the method of deterministic equivalents with finite-dimensional Monte Carlo simulations, in order to give an idea on the method accuracy. [6]

[6]Precise statements on the order of convergence with respect to M of the actual finite dimensional SINRs to their deterministic equivalents are given in [10].



(a) $r^\star = 6$ (b) $r^\star = 11$

Fig. 2. Comparison of sum spectral efficiency (bit/s/Hz) vs. SNR (dB) for JSDM with their corresponding deterministic equivalents. "JGP" denotes JSDM with *joint group processing* and "PGP" denotes JSDM with *per-group processing*.

In the following examples, the BS is equipped with a uniform circular array with $M = 100$ isotropic antenna elements equally spaced on a circle of radius λD, for $D = \frac{0.5}{\sqrt{(1-\cos(2\pi/M))^2+\sin(2\pi/M)^2}}$, resulting in the minimum distance between antenna elements equal to $\frac{\lambda}{2}$. Users form $G = 6$ symmetric groups, with AS $\Delta = 15^o$ and azimuth AoA $\theta_g = -\pi + \Delta + (g-1)\frac{2\pi}{G}$ for $g = 1, \ldots, G$. The user channel correlation is obtained according to (2). For the system geometry defined above, the transmit covariance matrix for each group has rank $r = 21$. However, half of the non-zero eigenvalues are extremely small, yielding an effective rank $r^\star = 11$. Somehow arbitrarily, we fixed to serve $S' = 5$ data streams per group, so that the total number of users being served is $S = S'G = 30$, and chose $b' = 10$.

Figs. 2(a) and 2(b) show the performance of the JSDM schemes when the pre-beamforming matrix is designed according to the approximate BD method described in Section IV-B, choosing $r^\star = 6$ and $r^\star = 12$, respectively. Given the noise unit variance normalization, we have that $\mathrm{SNR} = P$. The solid "squares" are obtained through simulations and the dotted "x" are obtained using the corresponding deterministic equivalent approximations. The regularization parameter is fixed to $\alpha = \frac{S}{bP}$ for both JGP and PGP. The performance of JSDM with JGP in Figs. 2(a) and 2(b) is identical, owing to the fact that we use eigen-beamforming with $\boldsymbol{B}_g = \boldsymbol{U}_g$, independent of $r^\star$. For the sake of comparison, the sum capacity of the MIMO BC channel with full CSIT (see (3)) is also shown (solid "circles" in green), obtained by the iterative waterfiling approach of [30].

Remark 5: By choosing $r^\star$ too small, such that significant eigenmodes are not taken into account by the approximate BD pre-beamforming matrix, the resulting inter-group interference is large and the performance of PGP is severely interference limited (e.g., Fig. 2(a)). Instead, by choosing $r^\star$ large enough, in order to include all significant eigenmodes, the performance of PGP does not show a noticeable interference limited behavior over a wide range of SNR. This is the case of Fig. 2(b), where we chose $r^\star = 12$ and the channel covariance matrix has rank $r = 21$, but only 11 significant eigenvalues. As a matter of fact, the PGP rate curves of Fig. 2(b) will eventually flatten, but this happens at extremely large SNR, irrelevant for practical applications. This example shows that $r^\star$ should always be chosen in order to include all strongest eigenmodes. However, making $r^\star = r$ is generally not a good choice since many eigenmodes may be very close to zero (as in this example) and therefore including them in the count of $r^\star$ yields a dimensionality bottleneck without any real benefit in terms of inter-group interference (recall that $r^\star G \leq M$, therefore if $r^\star$ is large we may have to decrease G, i.e., serve less groups in parallel). We conclude that the choice of the effective rank $r^\star$ should be carefully optimized, depending on the specific channel covariance eigenvalue distribution. $\lozenge$

VI. Downlink training and noisy CSIT

In this section, we evaluate the impact of noisy CSIT by including the fact that the effective channels are estimated by the UTs from the downlink training phase. In the vast literature dedicated to CSIT feedback (see for example [2] and references therein), methods that achieve the estimated channel Mean-Square Error (MSE) decreases as $O(1/P^\beta)$ for some $\beta \geq 1$, even in the presence of channel feedback noise and errors, are well-known. In contrast, the MSE due to estimation from the downlink training phase decreases at best as $O(1/P)$. In fact, this is given by the high-SNR behavior of the MMSE for a Gaussian signal (the channel vectors) in Gaussian noise. If the CSIT feedback scheme is designed to achieve exponent $\beta > 1$ and the channel SNR is sufficiently large, the feedback error is negligible with respect to the downlink estimation error [2]. Hence, for simplicity, we consider the optimistic situation of ideal and delay-free CSIT feedback, and focus only on the effect of the downlink channel estimation error and dimensionality penalty factor of the training phase (a similar approach is followed in [4]).

For brevity, we focus only on the case of PGP.[7] From Section V-B, the channel covariance matrix for a user g_k is given by $\bar{\boldsymbol{R}}_g = \boldsymbol{B}_g^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_g$. In order to estimate the effective channel vector $\mathbf{h}_{g_k}$, i.e., the column of the effective channel matrix $\mathbf{H}_g$ corresponding to user g_k, the BS sends unitary training sequences of length b', in parallel over the b' virtual inputs of the pre-beamforming of each group g. Hence, the training phase with PGP spans b' symbols. The UTs in each group make use of linear MMSE estimation, which is the optimal estimator for minimizing the

[7] Analogous results can be obtained for the case of JGP, but these are practically less interesting since JGP requires typically too large training and feedback overhead in FDD systems.

MSE since the observation at each user and the channel vector are conditionally jointly Gaussian given the training sequences. The MMSE channel estimates are fed back to the BS and are used to compute the linear precoders $\{\boldsymbol{P}_g\}$. Assuming that in each coherence block of T symbols the training phase makes use of b' symbols, and the remaining $T-b'$ symbols are available for downlink data transmission, it follows that the spectral efficiency must be scaled by the dimensionality penalty factor $\max\{1-b'/T, 0\}$.

We consider a scheme where a scaled unitary training matrix $\boldsymbol{X}_{\rm tr}$ of dimension $b' \times b'$ is sent, simultaneously, to all groups in the common downlink training phase. The corresponding received signal at group g receivers is given by

$$\boldsymbol{Y}_g = \mathsf{H}_g^{\mathsf{H}} \boldsymbol{X}_{\rm tr} + \sum_{g' \neq g} \boldsymbol{H}_g{}^{\mathsf{H}} \boldsymbol{B}_{g'} \boldsymbol{X}_{\rm tr} + \boldsymbol{Z}_g. \tag{47}$$

Multiplying from the right by $\boldsymbol{X}_{\rm tr}^{\mathsf{H}}$ and using the fact that, by design, $\boldsymbol{X}_{\rm tr}\boldsymbol{X}_{\rm tr}^{\mathsf{H}} = \rho_{\rm tr}\boldsymbol{I}_{b'}$ where $\rho_{\rm tr}$ is the power allocated to training, we obtain

$$\boldsymbol{Y}_g \boldsymbol{X}_{\rm tr}^{\mathsf{H}} = \rho_{\rm tr} \mathsf{H}_g^{\mathsf{H}} + \rho_{\rm tr} \sum_{g' \neq g} \boldsymbol{H}_g{}^{\mathsf{H}} \boldsymbol{B}_{g'} + \boldsymbol{Z}_g \boldsymbol{X}_{\rm tr}^{\mathsf{H}}. \tag{48}$$

Extracting the g_k-th row, dividing by $\sqrt{\rho_{\rm tr}}$, using the fact that $\boldsymbol{Z}_g\boldsymbol{X}_{\rm tr}^{\mathsf{H}}$ has i.i.d. entries $\sim \mathcal{CN}(0, \rho_{\rm tr})$ and taking Hermitian transpose of everything, we obtain the noisy observation for estimating the g_k-th effective channel vector in the form

$$\widetilde{\mathsf{h}}_{g_k} = \sqrt{\rho_{\rm tr}} \mathsf{h}_{g_k} + \sqrt{\rho_{\rm tr}} \left(\sum_{g' \neq g} \boldsymbol{B}_{g'}^{\mathsf{H}} \right) \boldsymbol{h}_{g_k} + \widetilde{\boldsymbol{z}}_{g_k}, \tag{49}$$

where $\widetilde{\boldsymbol{z}}_{g_k} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{b'})$. The MMSE estimator for h_{g_k} based on (49) is given by

$$\begin{aligned}
\widehat{\mathsf{h}}_{g_k} &= \mathbb{E}\left[\mathsf{h}_{g_k}\widetilde{\mathsf{h}}_{g_k}^{\mathsf{H}}\right] \mathbb{E}\left[\widetilde{\mathsf{h}}_{g_k}\widetilde{\mathsf{h}}_{g_k}^{\mathsf{H}}\right]^{-1} \widetilde{\mathsf{h}}_{g_k} \\
&= \sqrt{\rho_{\rm tr}} \left[\boldsymbol{B}_g^{\mathsf{H}} \boldsymbol{R}_g \sum_{g'=1}^{G} \boldsymbol{B}_{g'} \right] \left[\rho_{\rm tr} \sum_{g',g''=1}^{G} \boldsymbol{B}_{g'}^{\mathsf{H}} \boldsymbol{R}_g \boldsymbol{B}_{g''} + \boldsymbol{I}_{b'} \right]^{-1} \widetilde{\mathsf{h}}_{g_k} \\
&= \frac{1}{\sqrt{\rho_{\rm tr}}} \left(\boldsymbol{M}_g \tilde{\boldsymbol{R}}_g \boldsymbol{O}^{\mathsf{T}} \right) \left[\boldsymbol{O} \tilde{\boldsymbol{R}}_g \boldsymbol{O}^{\mathsf{T}} + \frac{1}{\rho_{\rm tr}} \boldsymbol{I}_{b'} \right]^{-1} \widetilde{\mathsf{h}}_{g_k}
\end{aligned} \tag{50}$$

where we used the fact that $\mathsf{h}_{g_k} = \boldsymbol{B}_g^{\mathsf{H}} \boldsymbol{h}_{g_k}$, where $\tilde{\boldsymbol{R}}_g$ is defined in (21) and we introduced the $b' \times b$ block matrices

$$\begin{aligned}
\boldsymbol{M}_g &= [\boldsymbol{0}, \ldots, \boldsymbol{0}, \underbrace{\boldsymbol{I}_{b'}}_{\text{block } g}, \boldsymbol{0}, \ldots, \boldsymbol{0}] \\
\boldsymbol{O} &= [\boldsymbol{I}_{b'}, \boldsymbol{I}_{b'}, \ldots, \boldsymbol{I}_{b'}].
\end{aligned}$$

Notice that in the case of perfect BD we have that $\boldsymbol{R}_g \boldsymbol{B}_{g'} = \boldsymbol{0}$ for $g' \neq g$. Therefore, (49) and (50) reduce to

$$\widetilde{\mathsf{h}}_{g_k} = \sqrt{\rho_{\rm tr}} \mathsf{h}_{g_k} + \widetilde{\boldsymbol{z}}_{g_k}, \tag{51}$$

and

$$\widehat{\mathbf{h}}_{g_k} = \frac{1}{\sqrt{\rho_{\rm tr}}}\bar{\boldsymbol{R}}_g \left[\bar{\boldsymbol{R}}_g + \frac{1}{\rho_{\rm tr}}\boldsymbol{I}_{b'}\right]^{-1} \widetilde{\mathbf{h}}_{g_k} \tag{52}$$

respectively, where we recall the definition $\bar{\boldsymbol{R}}_g = \boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{R}_g\boldsymbol{B}_g$.

For this channel estimation scheme, the deterministic equivalent approximation of the SINR terms for RZFBF and ZFBF precoding can be obtained following [10], [31], the approach of which can be directly applied to our case, and using the well-known MMSE decomposition

$$\mathbf{h}_{g_k} = \widehat{\mathbf{h}}_{g_k} + \widehat{\mathbf{e}}_{g_k}, \tag{53}$$

with $\mathbb{E}[\widehat{\mathbf{h}}_{g_k}\widehat{\mathbf{h}}_{g_k}^{\mathsf{H}}] = \widehat{\bar{\boldsymbol{R}}}_g$ and MMSE covariance matrix $\mathbb{E}[\widehat{\mathbf{e}}_{g_k}\widehat{\mathbf{e}}_{g_k}^{\mathsf{H}}] = \bar{\boldsymbol{R}}_g - \widehat{\bar{\boldsymbol{R}}}_g$. For completeness, the fixed-point equations leading to the deterministic equivalent SINR approximation for PGP with noisy CSIT are given in Appendix A. Eventually, the achievable rate of user g_k is approximated by

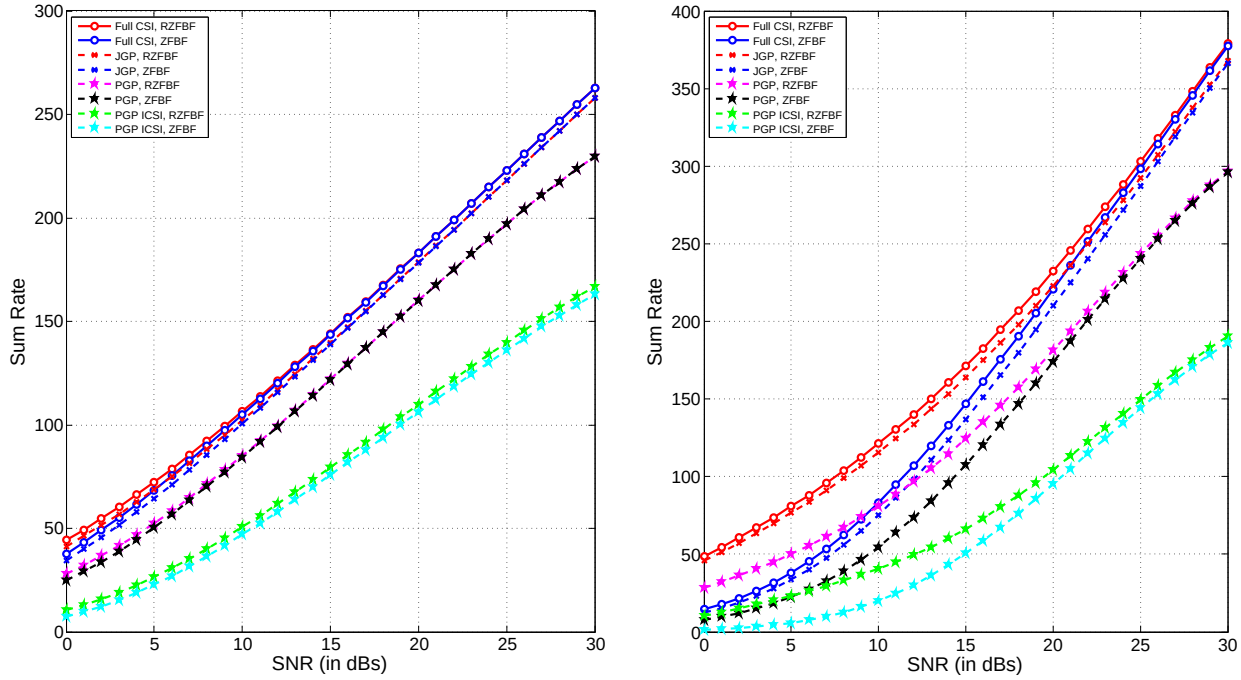
$$R_{g_k,\rm pgp,csit} \quad = \quad \max\left\{1 - \frac{b'}{T}, 0\right\} \times \log(1 + \widehat{\gamma}^{o}_{g_k,\rm pgp,csit}), \tag{54}$$

where $\widehat{\gamma}^{o}_{g_k,\rm pgp,csit}$ indicates either $\widehat{\gamma}^{o}_{g_k,\rm pgp,rzf,csit}$ or $\widehat{\gamma}^{o}_{g_k,\rm pgp,zf,csit}$, as detailed in Appendix A.

Remark 6: Assuming that, as $M \to \infty$, the other system dimensions $r^\star, S$ and b also go to infinity linearly with M, the achievable rate approximation error converges to zero almost surely as $M \to \infty$. However, the dimensionality factor $\max\{1 - b'/T, 0\}$ is equal to zero for $b' \geq T$. Hence, in order to obtain mathematically meaningful results we assume that also the coherence block length T grows linearly with M, and we define the factor $\tau = b'/T$ as the *dimensionality crowding factor* of the channel. In practice, this means that the method is valid in the regime of b' large, but still significantly smaller than T. $\lozenge$

A. Results with downlink channel estimation

We demonstrate the effect of noisy CSIT on the performance of RZFBF and ZFBF in Fig. 3, for the same antenna configuration of Section V-C with $r^\star = 11$, for $S' = 4$ and $S' = 8$ streams per group. For the sake of comparison, the solid "red" ("blue") curve denotes the sum spectral efficiency achieved by RZFBF (ZFBF) with full noiseless CSIT, i.e., by computing the precoding matrix in one step, directly from the instantaneous channel matrix $\underline{\boldsymbol{H}}$. The dotted lines represent the performance of JSDM for JGP with eigen-beamforming and noiseless CSIT (i.e., perfect knowledge of the effective channel $\underline{\mathbf{H}}$). The "magenta" ("black") curves denote the sum spectral efficiency for JSDM with PGP and approximate BD, also in the case of noiseless CSIT. Finally, the "green" ("cyan") curves denote the achievable sum spectral efficiency for JSDM with PGP and noisy CSIT, obtained by downlink training and MMSE estimation as explained above. These curves are obtained by optimizing the parameter b', for given S', $r^\star$ and SNR. Since a set of training sequences is sent simultaneously to all groups, the training power is given by $\rho_{\rm tr} = \frac{P}{G}$, such that the total sum power constraint is preserved also during the training phase.

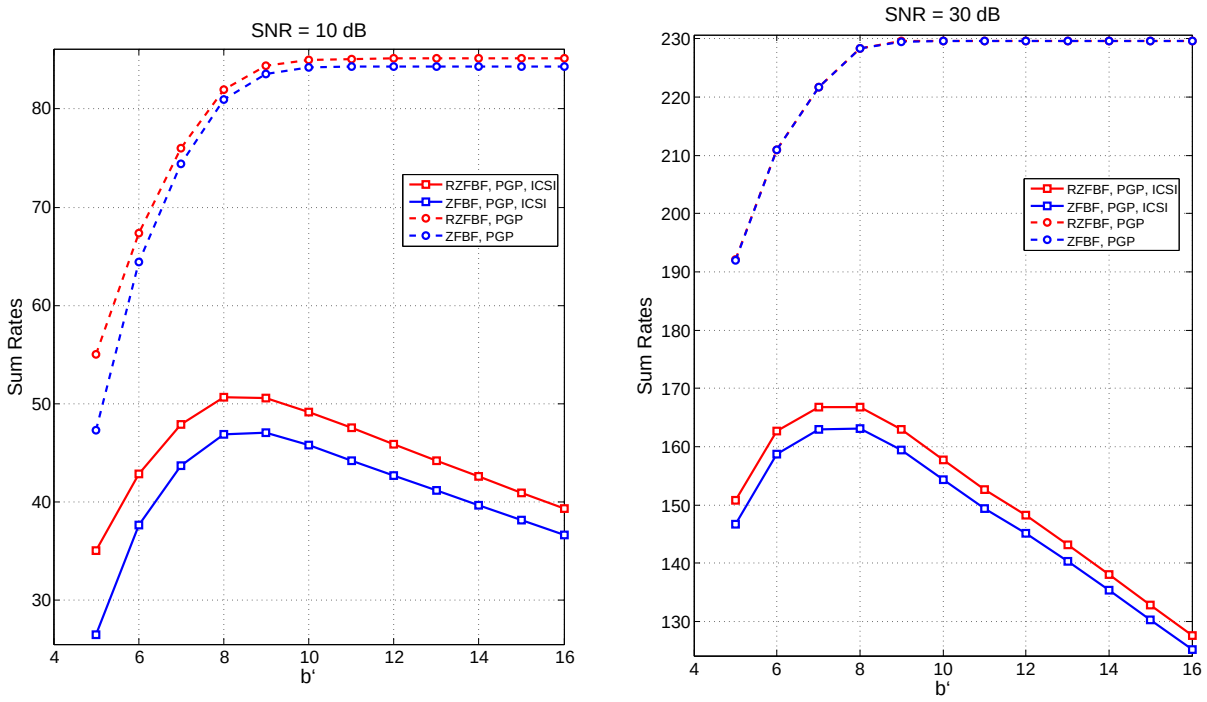


(a) $S' = 4$ (b) $S' = 8$

Fig. 3. Sum spectral efficiency (bit/s/Hz) vs. SNR (dB) for JSDM (computed via deterministic equivalents) with $r^\star = 11$, for $S' = 4$ and $S' = 8$. The coherence block length is $T = 40$. The "green" and "cyan" curves denote the results for imperfect CSIT with optimized choice of b'. "JGP" denotes JSDM with *joint group processing* and "PGP" denotes JSDM with *per-group processing*.

Remark 7: We examine now the optimization of the parameter b' for fixed target S', in the case of downlink training and noisy CSIT. Having fixed $r^\star$ as discussed in Remark 5, and assuming $0 \leq S' \leq b' \leq M - r^\star(G-1)$, for each value of SNR and given JSDM precoding scheme there is an optimal choice of b'. For example, Fig. 4 shows the dependency of the sum spectral efficiency of JSDM with PGP with respect to b' for $S' = 8$ and $\mathsf{SNR} = 10$ and 30 dB. We notice that the sum spectral efficiency including channel estimation is not monotonically increasing with b'. In fact, letting b' large yields better conditioned effective channel matrices, but incurs a larger dimensionality cost of the downlink training phase. The tension between these two issues yields a non-trivial choice for the optimal value of b' maximizing the system spectral efficiency. Similar trends can be observed for different values of S' and different values of SNR. $\lozenge$

Remark 8: Having chosen b', we focus now on choice of optimal S'. This depends heavily on the precoding scheme and the operating SNR. For a given operating SNR, there is approximately a linear dependence between the optimal S' and b' for both the RZFBF and the ZFBF precoders considered here. This linear dependence can be characterized by a single parameter, namely, the slope of the line relating the optimal S' and b'. In Fig. 5 we have plotted this slope versus SNR. It can be seen that for RZFBF, at low values of SNR, the choice $S' = b'$ (slope equal to 1) is optimal. In contrast, for ZFBF it is better to serve some $S' < b'$ number of users. As the SNR increases, the ZFBF slope increases and approaches that of the same slope of RZFBF at high SNR.



(a) $S' = 8$, SNR = 10 dB

(b) $S' = 8$, SNR = 30 dB

Fig. 4. Sum spectral efficiency (bit/s/Hz) vs. b' for JSDM with $r^\star = 11$, for $S' = 8$ (computed via deterministic equivalents). The coherence block length $T = 40$. The "dashed" curves denote the results for PGP with perfect CSIT, and the "solid" lines denote the same for imperfect CSIT.

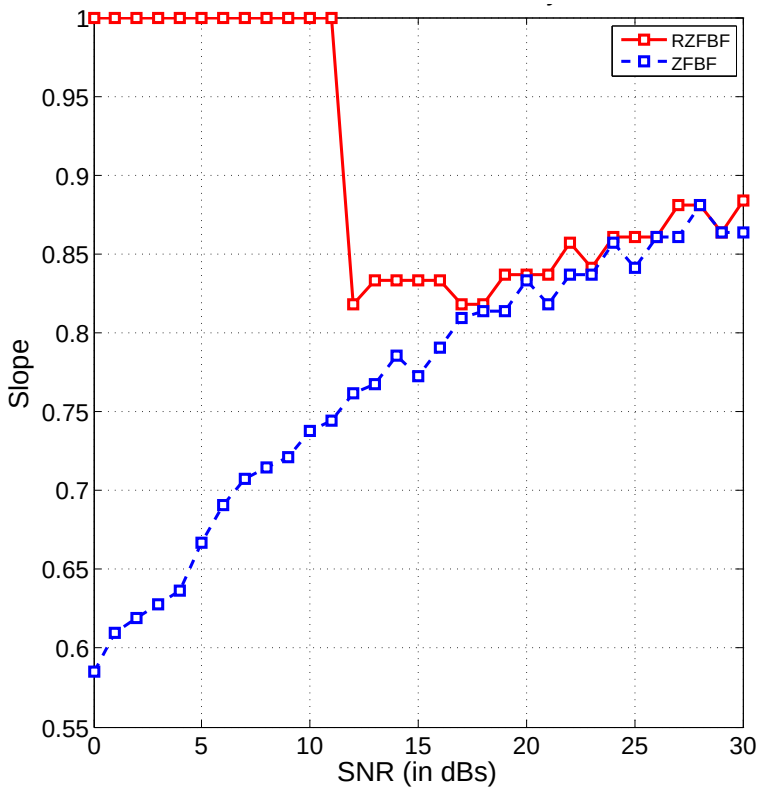


Fig. 5. Ratio S'/b' (slope) for the optimized S' and b' versus the channel SNR for different precoders.

VII. UNIFORM LINEAR ARRAYS: EIGENVALUES AND EIGENVECTORS

In this section we consider the antenna correlation model (2) for the special but important case of a Uniform Linear Array (ULA) of large dimension ($M \gg 1$), and obtain important insight on the behavior of the normalized asymptotic rank $\rho = \lim_{M\to\infty} \frac{r}{M}$ and of the eigenvectors $\boldsymbol{U}$ of the covariance matrix $\boldsymbol{R}$. We consider a 120 deg sector, obtained by using directional radiating elements, and assume that the sector is centered around the x-axis ($\alpha = 0$ azimuth angle), and that no energy is received for angles $\alpha \notin [-\pi/3, \pi/3]$. A ULA formed by M such directional radiating elements is placed at the origin along the y-axis. Denoting by λD the spacing of the antenna elements, the covariance matrix of the channel for a user at AoA θ and AS Δ according to the model of Section II is given by the Toeplitz form

$$[\boldsymbol{R}]_{m,p} = \frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} e^{-j2\pi D(m-p)\sin(\alpha)} d\alpha \tag{55}$$

for $m, p \in \{0, 1, \ldots, M-1\}$. In order to characterize eigenvalues and eigenvectors of $\boldsymbol{R}$ with respect to D, Δ, θ, for large M, we resort to the well-known results of [17], [32].

From [17], we recall the following fundamental result. Let $S(\xi)$ be a uniformly bounded absolutely integrable function over $\xi \in [-1/2, 1/2]$, i.e.,

$$\int_{-1/2}^{1/2} |S(\xi)| d\xi < \infty, \;\; \kappa_1 \leq S(\xi) \leq \kappa_2,$$

where the bounds hold for all $\xi \in [-1/2, 1/2]$ up to a set of measure zero. Assume that we can write the sequence $r_m = [\boldsymbol{R}]_{\ell,\ell-m}$ as the inverse discrete-time Fourier transform of $S(\xi)$, i.e.,

$$r_m = \int_{-1/2}^{1/2} S(\xi) e^{j2\pi\xi m} d\xi. \tag{56}$$

Then, the Toeplitz matrix $\boldsymbol{R}$ can be approximated by the circulant matrix $\boldsymbol{C}$ defined by its first column with m-th element

$$c_m = \begin{cases} r_m + r_{m-M} & \text{for } \; m = 1, \ldots, M-1 \\ r_0 & \text{for } \; m = 0 \end{cases}, \tag{57}$$

where the approximation holds in the following sense:

Fact 1: The set of eigenvalues $\{\lambda_m(\boldsymbol{R})\}$, $\{\lambda_m(\boldsymbol{C})\}$ and the set of uniformly spaced samples $\{S(m/M) : m = 0, \ldots, M-1\}$ are asymptotically *equally distributed*, i.e., for any continuous function $f(x)$ defined over $[\kappa_1, \kappa_2]$, we have

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} f(\lambda_m(\boldsymbol{R})) = \lim_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M-1} f(\lambda_m(\boldsymbol{C})) = \int_{-1/2}^{1/2} f(S(\xi)) d\xi. \tag{58}$$

■

Fact 2: The eigenvectors of $\boldsymbol{R}$ are approximated by the eigenvectors of $\boldsymbol{C}$ in the following eigenspace approximation sense. Define the asymptotic eigenvalue cumulative distribution function (CDF) of the eigenvalues of $\boldsymbol{R}$ to be the

right-continuous non-decreasing function $F(\lambda)$ such that $F(\lambda) = \int_{S(\xi)\leq\lambda} d\xi$ for any point of continuity $\kappa_1 \leq \lambda \leq \kappa_2$. Let $\lambda_0(\boldsymbol{R}) \leq \ldots, \leq \lambda_{M-1}(\boldsymbol{R})$ and $\lambda_0(\boldsymbol{C}) \leq \ldots, \leq \lambda_{M-1}(\boldsymbol{C})$ denote the set of ordered eigenvalues of $\boldsymbol{R}$ and $\boldsymbol{C}$, and let $\boldsymbol{U} = [\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{M-1}]$ and $\boldsymbol{F} = [\boldsymbol{f}_0, \ldots, \boldsymbol{f}_{M-1}]$ denote the corresponding eigenvectors. [8] For any interval $[a, b] \subseteq [\kappa_1, \kappa_2]$ such that $F(\lambda)$ is continuous on $[a, b]$, consider the eigenvalues index sets $\mathcal{I}_{[a,b]} = \{m : \lambda_m(\boldsymbol{R}) \in [a, b]\}$ and $\mathcal{J}_{[a,b]} = \{m : \lambda_m(\boldsymbol{C}) \in [a, b]\}$, and define $\boldsymbol{U}_{[a,b]} = (\boldsymbol{u}_m : m \in \mathcal{I}_{[a,b]})$ and $\boldsymbol{F}_{[a,b]} = (\boldsymbol{f}_m : m \in \mathcal{J}_{[a,b]})$ be the submatrices of $\boldsymbol{U}$ and $\boldsymbol{F}$ formed by the columns whose indices belong to the sets $\mathcal{I}_{[a,b]}$ and $\mathcal{J}_{[a,b]}$, respectively. Then, the eigenvectors of $\boldsymbol{C}$ approximate the eigenvectors of $\boldsymbol{R}$ in the sense that

$$\lim_{M\to\infty} \frac{1}{M} \left\| \boldsymbol{U}_{[a,b]} \boldsymbol{U}_{[a,b]}^{\mathsf{H}} - \boldsymbol{F}_{[a,b]} \boldsymbol{F}_{[a,b]}^{\mathsf{H}} \right\|_F^2 = 0. \tag{59}$$

■

A well-known property of circulant matrices [32] is that their eigenvectors form a unitary DFT matrix, i.e., the matrix whose (ℓ, m)-th element is given by $[\boldsymbol{F}]_{\ell,m} = \frac{e^{-j2\pi\ell m/M}}{\sqrt{M}}$. This has an important consequence for JSDM with large ULAs: in the regime of large M where the Toeplitz channel correlation matrix $\boldsymbol{R}$ is well approximated by its circulant version $\boldsymbol{C}$, we can approximate $\boldsymbol{U}$, the tall unitary matrix of the channel covariance eigenvectors, with a submatrix of $\boldsymbol{F}$, formed by a selection of columns of $\boldsymbol{F}$. Hence, we can design the pre-beamforming stage of JSDM by replacing $\boldsymbol{U}$ with its DFT approximation, avoiding the need of a precise estimation of the actual channel covariance matrix. In order to understand how to select the columns of $\boldsymbol{F}$, we need to gain more insight into the asymptotic behavior of the eigenvalues of $\boldsymbol{R}$.

For $r_m = [\boldsymbol{R}]_{\ell,\ell-m}$ with $[\boldsymbol{R}]_{m,p}$ given by (55), and $\boldsymbol{C}$ defined as in (57), the eigenvalues $\{\lambda_k(\boldsymbol{C})\}$ can be given explicitly for any finite M as follows:

$$\begin{aligned}
\lambda_k(\boldsymbol{C}) &= \sum_{m=0}^{M-1} c_m e^{-j\frac{2\pi}{M}mk} \\
&= r_0 + \sum_{m=1}^{M-1} [r_m + r_{m-M}] e^{-j\frac{2\pi}{M}mk} \\
&= r_0 + \sum_{m=1}^{M-1} r_m e^{-j\frac{2\pi}{M}mk} + \sum_{m=1}^{M-1} r_m^* e^{j\frac{2\pi}{M}mk} \\
&= \frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} \left[1 + 2\mathrm{Re}\left\{ \sum_{m=0}^{M-1} e^{-j2\pi m\omega_k(D,\alpha)} - 1 \right\} \right] d\alpha \\
&= -1 + \frac{1}{\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} \cos\left(\pi\omega_k(D,\alpha)(M-1)\right) \frac{\sin\left(\pi\omega_k(D,\alpha)M\right)}{\sin(\pi\omega_k(D,\alpha))} d\alpha,
\end{aligned} \tag{60}$$

[8] Notice that in the channel model defined in Section II we defined $\boldsymbol{U}$ of dimensions $M \times r$ to be the matrix of eigenvectors corresponding to the non-zero eigenvalues of $\boldsymbol{R}$. In the statement of this result, instead, $\boldsymbol{U}$ denotes the whole $M \times M$ matrix of eigenvectors, including the non-unique eigenvectors forming a unitary basis for the nullspace of $\boldsymbol{R}$, in the case $r < M$.

where we define the quantity $\omega_k(D,\alpha) = D\sin(\alpha) + k/M$.

In order to obtain the limiting CDF of the eigenvalues of $\boldsymbol{R}$ and find a simple formula for the asymptotic rank ρ, we obtain an explicit expression of $S(\xi)$ for the autocorrelation function $r_m = [\boldsymbol{R}]_{\ell,\ell-m}$. Using (55) and invoking the Lebesgue dominated convergence theorem, we have

$$\begin{aligned} S(\xi) &= \sum_{m=-\infty}^{\infty} r_m e^{-j2\pi\xi m} \\ &= \frac{1}{2\Delta}\int_{-\Delta+\theta}^{\Delta+\theta}\left[\sum_{m=-\infty}^{\infty} e^{-j2\pi m(D\sin(\alpha)+\xi)}\right] d\alpha \\ &\overset{(a)}{=} \frac{1}{2\Delta}\int_{-\Delta+\theta}^{\Delta+\theta}\left[\sum_{m=-\infty}^{\infty} \delta(D\sin(\alpha)+\xi-m)\right] d\alpha \\ &\overset{(b)}{=} \frac{1}{2\Delta}\int_{D\sin(-\Delta+\theta)}^{D\sin(\Delta+\theta)}\left[\sum_{m=-\infty}^{\infty} \delta(z+\xi-m)\right] \frac{dz}{\sqrt{D^2-z^2}}, \end{aligned} \tag{61}$$

where in (a) we used the Poisson sum formula (also known as "picked fence miracle" [33]), in (b) we made the change of variable $z = D\sin(\alpha)$. The expression (61) is valid for $-\frac{\pi}{2} \le \theta-\Delta < \theta+\Delta \le \frac{\pi}{2}$. A more general formula, able to recover the classical Bessel J_0 autocorrelation function [34] in the case of uniform isotropic scattering, is provided in Appendix B. Owing to the property of the Dirac delta function, we arrive at

$$S(\xi) = \frac{1}{2\Delta}\sum_{m\in[D\sin(-\Delta+\theta)+\xi,\, D\sin(\Delta+\theta)+\xi]} \frac{1}{\sqrt{D^2-(m-\xi)^2}}. \tag{62}$$

We have:

Lemma 1: The function $S(\xi)$ is non-constant over its support and uniformly bounded, provided that $D \in [0, 1/2]$ and $-\phi \le \theta-\Delta < \theta+\Delta \le \phi$ for some constant angle $\phi \in [0, \pi/2)$.

Proof: $S(\xi)$ is periodic and it is sufficient to restrict ξ to the interval $[-1/2, 1/2]$. As observed before, if $-\frac{\pi}{2} \le \theta-\Delta < \theta+\Delta \le \frac{\pi}{2}$, the general expression of $S(\xi)$ given in Appendix B coincides with (62), and we have $-D < -D\sin(\phi) \le D\sin(-\Delta+\theta) < D\sin(\Delta+\theta) \le D\sin(\phi) < D$. Since $-1/2 \le \xi \le 1/2$ and $D \in [0, 1/2]$, the following inequalities hold:

$$\begin{aligned} D\sin(-\Delta+\theta)+\xi &\ge D\sin(-\Delta+\theta) - 1/2 > -D - 1/2 \ge -1 \\ D\sin(\Delta+\theta)+\xi &\le D\sin(\Delta+\theta) + 1/2 < D + 1/2 \le 1. \end{aligned}$$

Since $-1 < D\sin(-\Delta+\theta)+\xi < D\sin(\Delta+\theta)+\xi < 1$, the only integer in the interval $[D\sin(-\Delta+\theta)+\xi, D\sin(\Delta+\theta)+\xi]$ is 0. Thus,

$$S(\xi) = \frac{1}{2\Delta}\sum_{0\in[D\sin(-\Delta+\theta)+\xi,\, D\sin(\Delta+\theta)+\xi]} \frac{1}{\sqrt{D^2-\xi^2}}. \tag{63}$$

The support $\mathcal{S}$ of $S(\xi)$ is the set of values $\xi \in [-1/2, 1/2]$ for which the interval $[D\sin(-\Delta+\theta)+\xi, D\sin(\Delta+\theta)+\xi]$ contains the point 0, i.e., $\mathcal{S} = [-D\sin(\Delta+\theta), -D\sin(-\Delta+\theta)]$. It is clear by inspection that $S(\xi)$ is not constant

over $\mathcal{S}$ (it is sufficient to observe that $S(\xi)$ is differentiable, and its derivative is not identically zero over a set of non-zero measure). To prove that $S(\xi)$ is uniformly bounded, it is sufficient to notice that the term $\frac{1}{\sqrt{D^2-\xi^2}}$ in (63) is real, continuous and finite for all $\xi \in (-D, D) \supset [-D\sin(\phi), D\sin(\phi)] \supseteq \mathcal{S}$. Hence, it attains its minimum κ' and maximum κ_2 on $\mathcal{S}$, and these are uniformly bounded as [9]

$$\frac{1}{D} \leq \kappa' < \kappa_2 \leq \frac{1}{D\cos(\phi)} < \infty.$$

■

Notice that the assumptions of Lemma 1 are satisfied for antenna spacing not larger than $\lambda/2$ and in the assumption, made here, that the ULA receives/transmits energy only in a 120 deg sector (i.e., for AoAs in $[-\pi/3, \pi/3]$). As a corollary of (62), we obtain the asymptotic rank in closed form:

Theorem 2: The asymptotic normalized rank of the channel covariance matrix $\boldsymbol{R}$ with elements defined in (55), with antenna separation λD, AoA θ and AS Δ, is given by

$$\rho = \min\{1, B(D, \theta, \Delta)\}, \tag{64}$$

where

$$B(D, \theta, \Delta) = \left| D\sin(-\Delta + \theta) - D\sin(\Delta + \theta) \right|. \tag{65}$$

Proof: Notice that $B(D, \theta, \Delta)$ is the size of the interval for m in the summation appearing in (62). If $B(D, \theta, \Delta) \geq 1$, for any $\xi \in [-1/2, 1/2]$ the sum in (62) is non-empty. It follows that $S(\xi) > 0$ for all ξ and the asymptotic normalized rank is $\rho = 1$. In contrast, if $B(D, \theta, \Delta) < 1$, there exist a set $\mathcal{S}^c \subseteq [-1/2, 1/2]$ of measure $1 - B(D, \theta, \Delta)$ for which if $\xi \in \mathcal{S}^c$ then the sum in (62) is empty. Therefore, in this case we have $\rho = B(D, \theta, \Delta)$. ■

A good approximation of the actual rank r for large but finite M is given by $r \approx \rho M$, where ρ is given by Theorem 2. Hence, we can predict accurately the rank of the channel covariance from the system geometric parameters (D, θ, Δ).

The empirical CDF of the eigenvalues of $\boldsymbol{R}$ is defined by

$$F_{\boldsymbol{R}}^{(M)}(\lambda) = \frac{1}{M} \sum_{m=1}^{M} 1\{\lambda_m(\boldsymbol{R}) \leq \lambda\}. \tag{66}$$

For large M, $F_{\boldsymbol{R}}^{(M)}(\lambda)$ can be approximated either using (60) or the collection of samples $\{S([m/M]) : m = 0, \ldots, M-1\}$, where $[x]$ indicates x modulo the interval $[-1/2, 1/2]$. In both cases, using the resulting collection of M values in (66), we obtain a convergent approximation $\widetilde{F}_{\boldsymbol{R}}^{(M)}(\lambda)$ of the empirical CDF (66) such that [17]

$$\lim_{M\to\infty} \widetilde{F}_{\boldsymbol{R}}^{(M)}(\lambda) = \lim_{M\to\infty} F_{\boldsymbol{R}}^{(M)}(\lambda) = F(\lambda).$$

[9] We use κ' instead of κ_1 to denote the minimum of $S(\xi)$ on its support since the minimum eigenvalue, denoted previously by κ_1, is generally equal to 0 whenever $\mathcal{S}$ is strictly included in $[-1/2, 1/2]$.

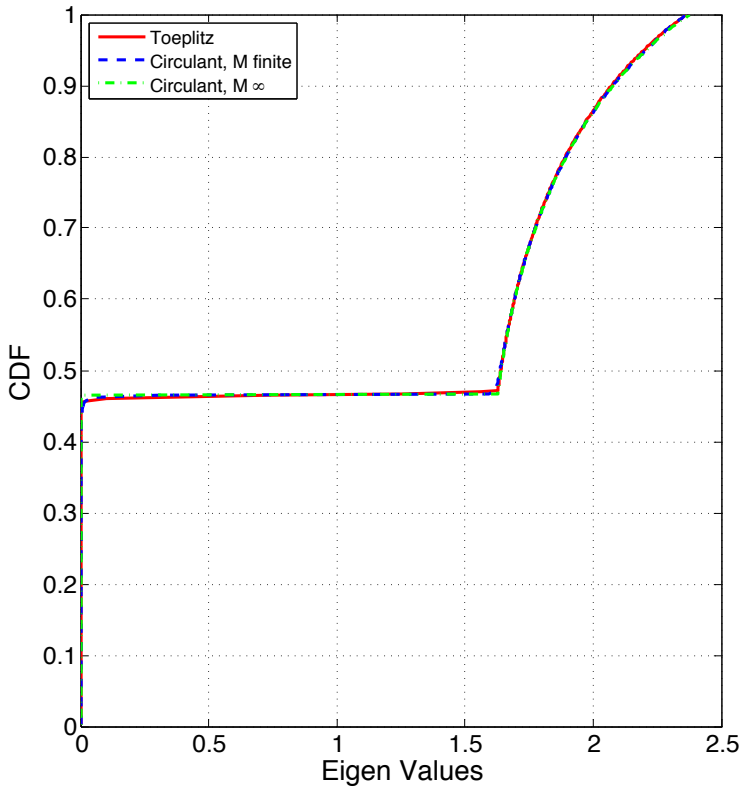


Fig. 6. $M = 400, \theta = \pi/6, D = 1, \Delta = \pi/10$. Exact empirical eigenvalue cdf of $\boldsymbol{R}$ (red), its approximation (60) based on the circulant matrix $\boldsymbol{C}$ (dashed blue) and its approximation from the samples of $S(\xi)$ (dashed green).

As an example, Fig. 6 shows the exact empirical CDF of $\boldsymbol{R}$, its circulant approximation obtained by (60)) and the asymptotic approximation obtained from the set $\{S([m/M]) : m = 0, \ldots, M-1\}$, for a specific choice of the system parameters. It is apparent that, in this regime, both approximations are very accurate.

A. Approximating the channel eigenspace

Going back to the problem of approximating the eigenvectors of $\boldsymbol{R}$ with a set of DFT columns, we notice the following properties of $S(\xi)$ in (62):

1) $S(\xi)$ has support on an interval $\mathcal{S} \subseteq [-1/2, 1/2]$, of length ρ (see proof of Theorem 2).
2) $S(\xi)$ is non-constant and bounded over its support (see Lemma 1).

It follows that $F(\lambda)$ has a single discontinuity at $\lambda = 0$, with jump of height $1 - \rho$, corresponding to the mass-point of the zero eigenvalues of $\boldsymbol{R}$. For $\rho < 1$, $F(\lambda)$ is continuous over $(0, \kappa_2]$ where $\kappa_2 = \max S(\xi) < \infty$ by Lemma 1. Hence, any interval $[a, b]$ with $0 < a < b \leq \kappa_2$ is a continuity interval of $F(\lambda)$, and the eigenspace approximation property of Fact 2 holds. In particular, we have established the following:

Corollary 1: Let $\mathcal{S}$ denote the support of $S(\xi)$, let $\mathcal{J}_{\mathcal{S}} = \{m : [m/M] \in \mathcal{S}, m = 0, \ldots, M-1\}$ be the set of indices for which the corresponding "angular frequency" $\xi_m = [m/M]$ belongs to $\mathcal{S}$, let $\boldsymbol{f}_m$ denote the m-th column of the unitary DFT matrix $\boldsymbol{F}$, and let $\boldsymbol{F}_{\mathcal{S}} = (\boldsymbol{f}_m : m \in \mathcal{J}_{\mathcal{S}})$ be the DFT submatrix containing the columns with indices in $\mathcal{J}_{\mathcal{S}}$. Then,

$$\lim_{M\to\infty} \frac{1}{M} \left\| \boldsymbol{U}\boldsymbol{U}^{\mathsf{H}} - \boldsymbol{F}_{\mathcal{S}}\boldsymbol{F}_{\mathcal{S}}^{\mathsf{H}} \right\|_F^2 = 0, \tag{67}$$

where $\boldsymbol{U}$ is the $M \times r$ "tall unitary" matrix of the non-zero eigenvectors of $\boldsymbol{R}$.

Proof: Since $S(\xi)$ is uniformly bounded and strictly positive over $\mathcal{S}$, we have $0 < \min_{\xi\in\mathcal{S}} S(\xi) = \kappa' < \max S(\xi) = \kappa_2$. Hence, letting $a = \kappa'$ and $b = \kappa_2$, and using the eigenspace approximation property of Fact 2 yields the result. ■

Consider now a JSDM configuration with an ULA serving G groups with AoAs within a 120 deg sector. For each group g, we can approximate the eigenmodes $\boldsymbol{U}_g$ by the DFT submatrix $\boldsymbol{F}_{\mathcal{S}_g}$, where $\mathcal{S}_g$ denotes the support of $S_g(\xi)$, given by (62) for AoA θ_g and AS Δ (for simplicity we assume that the AS is common to all groups, although this can be easily generalized). Corollary (1) implies that if $\mathcal{S}_g \cap \mathcal{S}_{g'} = \emptyset$ (disjoint angular frequency support), then $\boldsymbol{F}_{\mathcal{S}_g}^{\mathsf{H}}\boldsymbol{F}_{\mathcal{S}_g'} = \boldsymbol{0}$. It follows that if the G groups are chosen to have spectra with disjoint support, then $[\boldsymbol{F}_{\mathcal{S}_1}, \ldots, \boldsymbol{F}_{\mathcal{S}_G}]$ is *exactly* tall unitary and, because of Fact 2, $\underline{\boldsymbol{U}} = [\boldsymbol{U}_1, \ldots, \boldsymbol{U}_G]$ is approximately tall unitary, for large M. The following result provides such condition expressed directly in terms of the AoA intervals.

Theorem 3: Groups g and g' with angle of arrival θ_g and $\theta_{g'}$ and common angular spread Δ have spectra with disjoint support if their AoA intervals $[\theta_g - \Delta, \theta_g + \Delta]$ and $[\theta_{g'} - \Delta, \theta_{g'} + \Delta]$ are disjoint.

Proof: Define:

$$\begin{aligned}
A_g &= \max(D\sin(\theta_g + \Delta), D\sin(\theta_g - \Delta)) \\
B_g &= \min(D\sin(\theta_g + \Delta), D\sin(\theta_g - \Delta)) \\
A_{g'} &= \max(D\sin(\theta_{g'} + \Delta), D\sin(\theta_{g'} - \Delta)) \\
B_{g'} &= \min(D\sin(\theta_{g'} + \Delta), D\sin(\theta_{g'} - \Delta)).
\end{aligned}$$

From (62) we notice that $S_g(\xi)$ and $S_{g'}(\xi)$ have disjoint supports if $A_g \le B_{g'}$ or $A_{g'} \le B_g$. Since the mapping $x \mapsto \sin(x)$ is one-to-one in the interval $[-\pi/3, \pi/3]$, this condition corresponds to $[\theta_g - \Delta, \theta_g + \Delta] \cap [\theta_{g'} - \Delta, \theta_{g'} + \Delta] = \emptyset$.

■

B. DFT pre-beamforming

Owing to the asymptotic eigenspace approximation and mutual orthogonality of the previous section, an efficient approach to JSDM design when the BS is equipped with a large ULA per sector consists of selecting groups of users with (almost) identical AoA intervals, and find G groups of such users with non-overlapping AoA intervals.

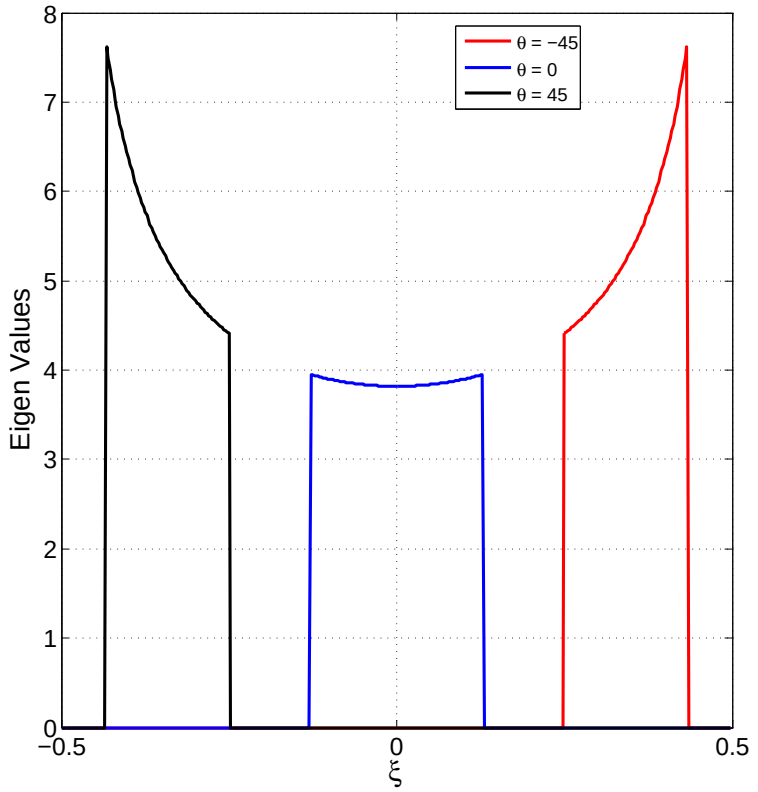


Fig. 7. Eigenvalue spectra for a ULA with $M = 400$, $G = 3$, $\theta_1 = \frac{-\pi}{4}, \theta_2 = 0, \theta_3 = \frac{\pi}{4}$, $D = 1/2$ and $\Delta = 15$ deg.

Then, we let $\boldsymbol{B}_g = \boldsymbol{F}_{\mathcal{S}_g}$, for $g = 1, \ldots, G$, with $\boldsymbol{F}_{\mathcal{S}_g}$ defined as in Corollary 1. It follows that $\boldsymbol{F}_{\mathcal{S}_g}^{\mathsf{H}} \boldsymbol{U}_{g'} \approx \boldsymbol{0}$ for all $g \neq g'$, such that the sum spectral efficiency achieved by JSDM with PGP is close to the sum spectral efficiency of the corresponding MU-MIMO downlink channel with full CSIT (see Theorem 1). Notice that this approach is particularly attractive since only a coarse parametric knowledge (AoA interval) for each user is required, rather than an accurate estimate of its channel covariance matrix.

Fig. 7 show the spectra $S_g(\xi)$ for $g = 1, 2, 3$, $M = 400$, and $\theta_1 = \frac{-\pi}{4}, \theta_2 = 0, \theta_3 = \frac{\pi}{4}$, with $D = 1/2$ and $\Delta = 15$ deg. The performance of JSDM with PGP and DFT pre-beamforming is shown in Fig. 8, indicating that up to 20 dB of SNR, DFT pre-beamforming performs close to schemes with full CSIT.

VIII. JSDM WITH 3D PRE-BEAMFORMING

So far we considered a planar geometry where each user group g is identified by its AoA interval $[\theta_g - \Delta, \theta_g + \Delta]$. For the sake of simplicity, we allocated equal power to all S downlink data streams. This is a near-optimal power allocation in the high SNR (high spectral efficiency) regime and in the case where the pathloss from the BS to all the UTs is approximately equal. In practice, however, users with same (or very similar) AoA interval may be located at different distances to the BS. In this case, a simple alternative to the complicated and generally non-convex power

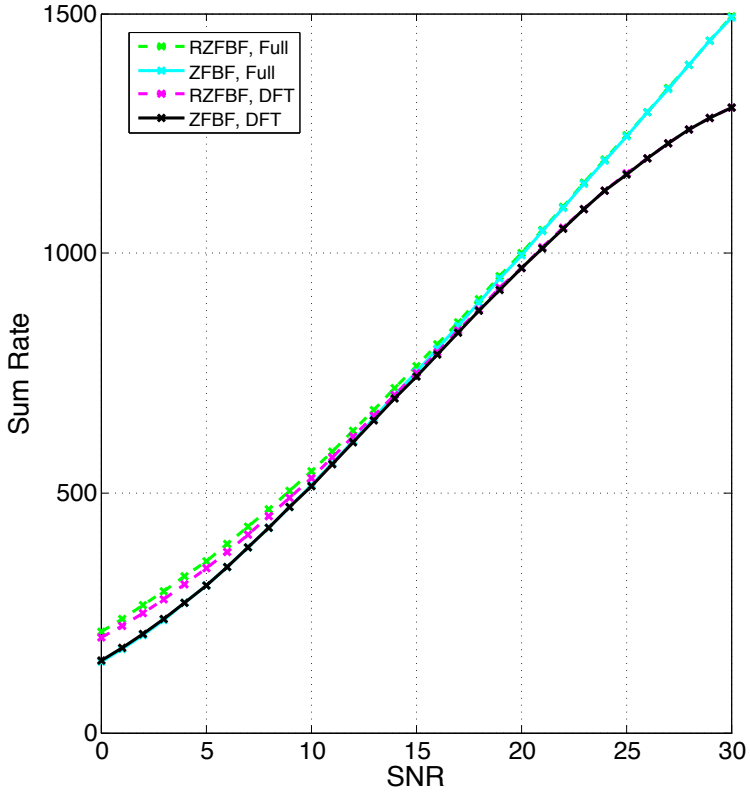


Fig. 8. Sum spectral efficiency (bit/s/Hz) vs. SNR (dB) for JSDM (computed via deterministic equivalents) for DFT pre-beamforming and PGP, for the configuration with spectra shown in Fig. 7, choosing $b_g = r_g$ for all groups $g = 1, 2, 3$.

allocation optimization across different users [10] consists of dividing the cell into concentric annular regions, and serve simultaneously groups in the same region, such that the pathloss is nearly equal for all jointly processed groups. Groups in different annular regions can be scheduled over the time-frequency slots. In this section, we consider an extension of this approach where we assume that the BS is elevated with respect to ground. For example, antenna elements could be placed on the window frames of a tall building forming a rectangular array with M antennas in each row (each row is an ULA) and a total of N rows in the vertical dimension. By exploiting the vertical dimension, different annular regions can be served simultaneously in the spatial domain.

Assuming a rectangular $N \times M$ array, we consider using a *separable* 3D pre-beamforming scheme: beamforming in the elevation angle dimension is used to form beams that "look down" at different angles, i.e., they illuminate concentric annular regions within the cell sector. For each such region, precoding in the azimuth angle dimension is

[10]While for MU-MIMO with full CSIT and optimal capacity achieving coding [23] the power allocation is a convex optimization problem that can be efficiently solved [35], for JSDM with either JGP and PGP, the problem is non-convex and the optimization is not amenable to a computationally efficient solution.

obtained by JSDM scheme with M antennas, as done before. Thanks to separability, we can optimize JSDM schemes independently, one for each annular region.

The groups served simultaneously by JSDM in the same region are now identified by two indices, (l, g) where $l = 1, \ldots, L$ indicates the annular region and $g = 1, \ldots, G_l$ the group in each l-th region. A set of groups served simultaneously, on the same time-frequency dimensions, is referred to as a "pattern". A pattern does not necessarily cover the whole sector. In fact, it is usually better to allow for "holes" in the pattern, i.e., the group footprints can be separated by gaps, in order to guarantee near orthogonality between the dominant eigenmodes of the groups in the same pattern and thus limiting inter-group interference with PGP. In order to provide coverage to the whole sector, different *intertwined patterns* can be multiplexed over the time-frequency dimension, similarly to the intertwined cooperative pattern idea proposed in [36], [37], [38]. The fraction of the time-frequency dimensions allocated to each pattern can be further optimized in order to maximize a network utility function, reflecting some desired notion of fairness (see for example [9]).

For the time being, we focus on a single pattern comprising L regions in the elevation angle dimension, and G_l groups in the azimuth angle dimension for each region $l = 1, \ldots, L$. We let $K_{l,g}$ denote the number of users in group (l, g). At the BS, an $N \times M$ rectangular antenna array with N rows and M columns is used. For each region l, we denote by $\boldsymbol{R}_{V,l} \in \mathbb{C}^{N \times N}$ the vertical channel covariance matrix [11] and, for each group (l, g), we let $\boldsymbol{R}_{H,l,g} \in \mathbb{C}^{M \times M}$ denote the the horizontal channel covariance matrix. $\boldsymbol{R}_{V,l}$ and $\boldsymbol{R}_{H,l,g}$ are modeled according to (2), with the eigen-decompositions:

$$\boldsymbol{R}_{V,l} = \boldsymbol{U}_{V,l} \boldsymbol{\Lambda}_{V,l} \boldsymbol{U}_{V,l}^{\mathsf{H}}, \quad \text{and} \quad \boldsymbol{R}_{H,l,g} = \boldsymbol{U}_{H,l,g} \boldsymbol{\Lambda}_{H,l,g} \boldsymbol{U}_{H,l,g}^{\mathsf{H}}. \tag{68}$$

Letting $\boldsymbol{h}_{l,g_k}$ denote the $MN \times 1$ the vectorized channel from the $M \times N$ BS array to the k^{th} user in group (l, g), we have

$$\mathbb{E}[\boldsymbol{h}_{l,g_k} \boldsymbol{h}_{l,g_k}^{\mathsf{H}}] = \boldsymbol{R}_{l,g} = \boldsymbol{R}_{H,l,g} \otimes \boldsymbol{R}_{V,l} = (\boldsymbol{U}_{H,l,g} \otimes \boldsymbol{U}_{V,l})(\boldsymbol{\Lambda}_{H,l,g} \otimes \boldsymbol{\Lambda}_{V,l})(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \otimes \boldsymbol{U}_{V,l}^{\mathsf{H}}). \tag{69}$$

This covariance matrix is common (by assumption) to all users g_k in group (l, g). Denoting the ranks of $\boldsymbol{R}_{H,l,g}$ and $\boldsymbol{R}_{V,l}$ by $r_{H,l,g}$ and $r_{V,l}$, respectively, we write $\boldsymbol{h}_{l,g_k}$ as

$$\boldsymbol{h}_{l,g_k} = (\boldsymbol{U}_{H,l,g} \otimes \boldsymbol{U}_{V,l})(\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}}) \boldsymbol{w}_{l,g_k},$$

where $\boldsymbol{U}_{H,l,g}$ is $M \times r_{H,l,g}$, $\boldsymbol{U}_{V,l}$ is $N \times r_{V,l}$, $\boldsymbol{\Lambda}_{H,l,g}$ is $r_{H,l,g} \times r_{H,l,g}$ and $\boldsymbol{\Lambda}_{V,l}$ is $r_{V,l} \times r_{V,l}$. The vector $\boldsymbol{w}_{l,g_k}$, of dimension $r_{H,l,g} r_{V,l} \times 1$, has i.i.d. entries $\sim \mathcal{CN}(0, 1)$.

In JSDM with 3D pre-beamforming, the transmitted signal is given by

$$\boldsymbol{x} = \sum_{l=1}^{L} (\boldsymbol{B}_l \boldsymbol{P}_l \boldsymbol{d}_l) \otimes \boldsymbol{q}_l, \tag{70}$$

[11] We assume that the vertical correlation does not depend on g, but just on l.

where $\boldsymbol{q}_l$ denotes the $N \times 1$ pre-beamforming vector for region l in the elevation angle dimension, $\boldsymbol{B}_l$ is the $M \times b_l$ pre-beamforming matrix of the form $\boldsymbol{B}_l = [\boldsymbol{B}_{l,1}, \ldots, \boldsymbol{B}_{l,G_l}]$, where $\boldsymbol{B}_{l,g}$ denotes the pre-beamforming matrix of size $M \times b_{l,g}$ for group (l,g) and $\boldsymbol{P}_l$ is the linear precoding matrix for the groups of region l, that depends on the instantaneous effective channels as given in Section III. Notice that we allocate (by design) a single dimension per region in the elevation angle direction (this is why $\boldsymbol{q}_l$ has dimensions $N \times 1$) since, because of the relatively small angle under which the BS sees the different regions, it is realistic to expect that $\boldsymbol{R}_{V,l}$ has a single dominant eigenmode. Generalizations considering higher dimensional vertical pre-beamforming for each region are conceptually straightforward, although not very useful in typical practical scenarios.

Using repeatedly the Kronecker product rule $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$, the received signal for user g_k in group (l,g) can be written as

$$\begin{aligned} y_{l,g_k} &= \boldsymbol{w}_{l,g_k}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}})(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \otimes \boldsymbol{U}_{V,l}^{\mathsf{H}})\boldsymbol{x} + z_{l,g_k} \\ &= \boldsymbol{w}_{l,g_k}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}})(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \otimes \boldsymbol{U}_{V,l}^{\mathsf{H}}) \left[\sum_{m=1}^{L} (\boldsymbol{B}_m \boldsymbol{P}_m \boldsymbol{d}_m) \otimes \boldsymbol{q}_m \right] + z_{l,g_k} \\ &= \boldsymbol{w}_{l,g_k}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}}) \sum_{m=1}^{L} \left[(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \boldsymbol{B}_m \boldsymbol{P}_m \boldsymbol{d}_m) \otimes (\boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_m) \right] + z_{l,g_k} \\ &= \boldsymbol{w}_{l,g_k}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}}) \sum_{m=1}^{L} \left[(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \boldsymbol{B}_m) \otimes (\boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_m) \right] \boldsymbol{P}_m \boldsymbol{d}_m + z_{l,g_k}. \end{aligned} \tag{71}$$

If $\boldsymbol{q}_m$ is chosen to be orthogonal to $\text{Span}(\{\boldsymbol{U}_{V,l} : l \neq m\})$, (71) reduces to

$$y_{l,g_k} = \boldsymbol{w}_{l,g_k}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}}) \left[(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \boldsymbol{B}_l) \otimes (\boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_l) \right] \boldsymbol{P}_l \boldsymbol{d}_l + z_{l,g_k}. \tag{72}$$

Stacking the signals y_{l,g_k} for all users g_k in group (l,g) into a $K_{l,g} \times 1$ vector $\boldsymbol{y}_{l,g}$, we obtain

$$\boldsymbol{y}_{l,g} = \boldsymbol{W}_{l,g}^{\mathsf{H}} (\boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}}) \left[(\boldsymbol{U}_{H,l,g}^{\mathsf{H}} \boldsymbol{B}_l) \otimes (\boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_l) \right] \boldsymbol{P}_l \boldsymbol{d}_l + \boldsymbol{z}_{l,g}, \tag{73}$$

where we let $\boldsymbol{W}_{l,g} = [\boldsymbol{w}_{l,g_1}, \ldots, \boldsymbol{w}_{l,g_{K_{l,g}}}]$ and $\boldsymbol{z}_{l,g} = [z_{l,g_1}, \ldots, z_{l,g_{K_{l,g}}}]^{\mathsf{T}}$.

If the regions are sufficiently separated in the elevation angle dimension, it is possible to align $\boldsymbol{q}_l$ with the dominant eigenmode of $\boldsymbol{U}_{V,l}$, while maintaining the orthogonality condition $\boldsymbol{U}_{V,m}^{\mathsf{H}} \boldsymbol{q}_l = 0$ for $m \neq l$. In this case, we have $\boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_l = (1, 0, \ldots, 0)^{\mathsf{T}}$ and (73) reduces to the same form treated previously for the planar geometry, with an additional region-specific coefficient $\sqrt{\lambda_{V,1}}$, corresponding to the largest eigenvalue of the matrix $\boldsymbol{\Lambda}_{V,l}$:

$$\boldsymbol{y}_{l,g} = \sqrt{\lambda_{V,1}} \boldsymbol{W}_{l,g}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,g}^{\frac{1}{2}} \boldsymbol{U}_{H,l,g}^{\mathsf{H}} \boldsymbol{B}_l \boldsymbol{P}_l \boldsymbol{d}_l + \boldsymbol{z}_{l,g}. \tag{74}$$

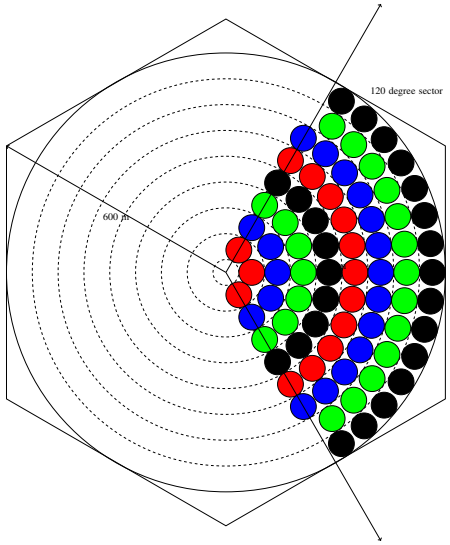


Fig. 9. The layout of one pattern for JSDM with 3D pre-beamforming. The concentric regions are separated by the vertical pre-beamforming. The circles indicate user groups. Same-color groups are served simultaneously using JSDM.

Stacking the vectors $\boldsymbol{y}_{l,g}$ for all $g = 1, \ldots, G_l$, we obtain

$$\boldsymbol{y}_l = \sqrt{\lambda_{V,1}} \begin{bmatrix} \boldsymbol{W}_{l,1}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,1}^{\frac{1}{2}} \boldsymbol{U}_{H,l,1}^{\mathsf{H}} \\ \boldsymbol{W}_{l,2}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,2}^{\frac{1}{2}} \boldsymbol{U}_{H,l,2}^{\mathsf{H}} \\ \vdots \\ \boldsymbol{W}_{l,G_l}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,G_l}^{\frac{1}{2}} \boldsymbol{U}_{H,l,G_l}^{\mathsf{H}} \end{bmatrix} \boldsymbol{B}_l \boldsymbol{P}_l \boldsymbol{d}_l + \boldsymbol{z}_l, \tag{75}$$

which is of the same form as (4). At this point, the pre-beamforming matrix $\boldsymbol{B}_l$ and the precoding matrix $\boldsymbol{P}_l$ can be optimized independently for each region l, as described before for the planar geometry. The coefficients $\lambda_{V,1}$ incorporate the effect of the different geometry of the annular regions in the elevation angle dimension, including the path loss due to different distances of the regions from the BS. The allocation of the total transmit power over the regions can be further optimized.

A. Results with 3D pre-beamforming

We present some results for JSDM with 3D pre-beamforming and PGP, with either BD or DFT pre-beamforming in each region. The system layout is shown in Fig. 9. We consider one sector of a hexagonal cell of radius 600 m. The scattering rings in the channel correlation model have radius $\mathsf{r} = 30$ m. The BS is located at the center of the cell with the antennas at an elevation of $\mathsf{h} = 50$ m, and is equipped with a rectangular array with $M = 200$ and $N = 300$. We partition the sector into 8 concentric regions at distance $60l$ m, $l \in \{1, \ldots, 8\}$. Each annular region is divided into small scattering rings, each defining a group. The pathloss between the BS and a point at distance x m

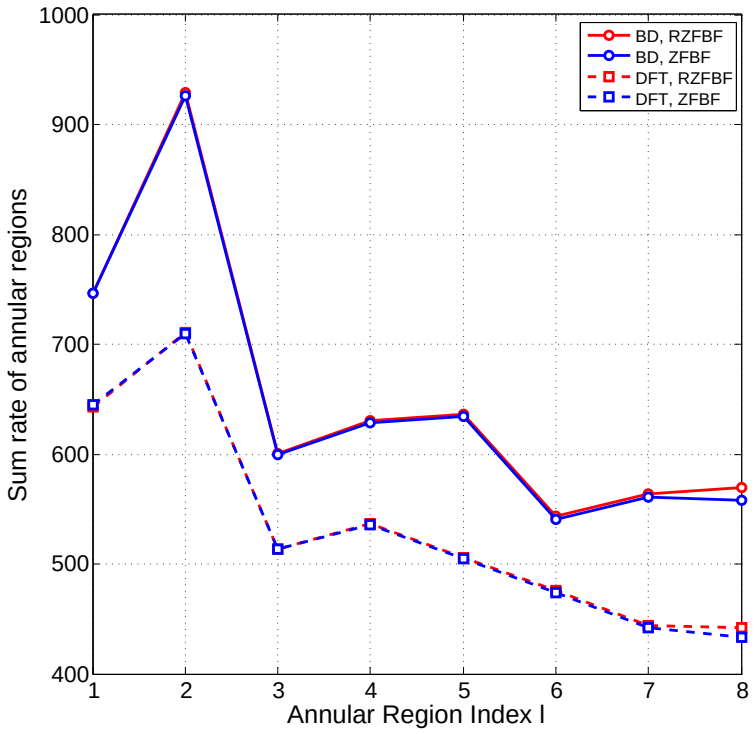


Fig. 10. Sum spectral efficiency $\bar{\mathcal{R}}_l$ for different annular regions $l = 1, \ldots, 8$ with Regularized ZF and ZF for JSDM with 3D pre-beamforming and ideal CSIT. "BD" denotes PGP with approximate block diagonalization and "DFT" stands for PGP with DFT pre-beamforming. Equal power is allocated to all served users and the number of users (streams) in each group is optimized in order to maximize the overall spectral efficiency.

is given by

$$g(\mathsf{x}) = \frac{1}{1 + (\frac{\mathsf{x}}{\mathsf{d}_0})^{\delta}}, \tag{76}$$

with $\delta = 3.8$, $\mathsf{d}_0 = 30$ m. In these results we assume ideal CSIT for computing the JSDM precoder. The horizontal covariance matrix for all groups (l, g) is given by (2) with $\Delta_{H,l} = \arctan(\frac{r}{60l}) = \arctan(\frac{1}{2l})$ and $\theta_{H,l,g} \in [-\pi/3, \pi/3]$ such that for any two groups (l, g_1) and (l, g_2), we have $|\theta_{H,l,g_1} - \theta_{H,l,g_2}| > 2\Delta_l$. It is easy to see from Fig. 9 that as the distance of the concentric regions from the BS increases, more and more user groups can be accommodated in the annular region, since $\Delta_{H,l}$ decreases. The vertical covariance matrix is again given by (2) with $\Delta_{V,l} = \frac{1}{2}(\arctan(\frac{60l+\mathsf{r}}{\mathsf{h}}) - \arctan(\frac{60l-\mathsf{r}}{\mathsf{h}}))$ and $\theta_{V,l} = \frac{1}{2}(\arctan(\frac{60l+\mathsf{r}}{\mathsf{h}}) + \arctan(\frac{60l-\mathsf{r}}{\mathsf{h}}))$. Since the total angle under which the sector is seen from the elevation viewpoint is narrow, a large number of antennas in the vertical direction is required in order to achieve orthogonality between all annular regions eigenmodes.

For finite N, in order to guarantee a desired angular separation between annular regions and therefore have near-orthogonality in the elevation angle dimension, it is convenient to partition the annular regions into maximally separated subsets (patterns) and apply BD in the vertical pre-beamforming. Different patterns can be scheduled in different time-

frequency slots. We denote by $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$ the set of patterns. Finding the best possible pattern partition is computationally hard, so for the sake of simplicity, we consider a simple partitioning as shown in Fig. 9, where annular regions with the same color belong to the same pattern. In our example (see Fig. 9), numbering the annular regions in ascending order based on their proximity to the BS, we have $\mathcal{A} = \{\{1,5\},\{2,6\},\{3,7\},\{4,8\}\}$. In this way, (75) is replaced by

$$\boldsymbol{y}_l = |\boldsymbol{\Lambda}_{V,l}^{\frac{1}{2}} \boldsymbol{U}_{V,l}^{\mathsf{H}} \boldsymbol{q}_l| \begin{bmatrix} \boldsymbol{W}_{l,1}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,1}^{\frac{1}{2}} \boldsymbol{U}_{H,l,1}^{\mathsf{H}} \\ \boldsymbol{W}_{l,2}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,2}^{\frac{1}{2}} \boldsymbol{U}_{H,l,2}^{\mathsf{H}} \\ \vdots \\ \boldsymbol{W}_{l,G_l}^{\mathsf{H}} \boldsymbol{\Lambda}_{H,l,G_l}^{\frac{1}{2}} \boldsymbol{U}_{H,l,G_l}^{\mathsf{H}} \end{bmatrix} \boldsymbol{B}_l \boldsymbol{P}_l \boldsymbol{d}_l + \boldsymbol{z}_l. \tag{77}$$

Notice that due to BD in the vertical direction, the inter-region interference is exactly zero since $\boldsymbol{U}_{V,m}^{\mathsf{H}} \boldsymbol{q}_l = \boldsymbol{0}$ for $m \neq l$. Within each annular region, we use JSDM with PGP. The pre-beamforming matrices $\boldsymbol{B}_l$ for region l are obtained using approximate BD or the DFT method, as discussed in previous sections. The dominant rank $r_{l,g}^{\star}$ for each group (l,g) is given by

$$r_{l,g}^{\star} = MD(\sin(\theta_{H,l,g} + \Delta_{H,l}) - \sin(\theta_{H,l,g} - \Delta_{H,l})), \tag{78}$$

which is a good approximation for large M motivated by Theorem 2. For simplicity, we do not consider noisy CSIT and assume that the BS has full knowledge of the effective channels. Hence, we let $b_{l,g} = r_{l,g}^{\star}$. In contrast, in the case of noisy CSIT the parameter $b_{l,g}$ should be optimized for given channel coherence block length T, as discussed in Remark 7.

Denoting by $\mathcal{R}_q$ the sum spectral efficiency of pattern $\mathcal{A}_q$, and letting Q denote the number of patterns, the network utility maximization problem is given by

$$\begin{aligned} \max \quad & g(\mathcal{R}_1, \ldots, \mathcal{R}_Q) \\ \text{subject to} \quad & \mathcal{R}_q \leq \nu_q \mathcal{R}_q^*, \quad \text{for } q = 1, \ldots, Q, \\ & \sum_{q=1}^{Q} \nu_q = 1. \end{aligned} \tag{79}$$

where $g(\cdot)$ is a concave component-wise non-decreasing network utility function capturing some desired notion of *fairness*, and the optimization variables $\{\nu_q\}$ are the fractions of time-frequency dimensions allocated to each pattern. We define

$$\mathcal{R}_q^* = \sum_{l \in \mathcal{A}_q} \sum_{g=1}^{G_l} \sum_{k=1}^{S_{l,g}} R_{l,g,k},$$

to be the spectral efficiency of each individual pattern, where $S_{l,g}$ is the number of downlink streams to group (l,g) and $R_{l,g,k}$ is the rate of the k-th stream of group (l,g). We have considered two cases of fairness: proportional fairness (PFS), and max-min fairness. In both cases, the optimal dimension allocation fractions $\{\nu_q\}$ can be found in closed

TABLE I

SUM SPECTRAL EFFICIENCY (BIT/S/HZ) UNDER PFS AND MAX-MIN FAIRNESS SCHEDULING FOR PGP AND APPROXIMATE BD/DFT.

Scheme	Approximate BD	DFT based
PFS, RZFBF	1304.4611	1067.9604
PFS, ZFBF	1298.7944	1064.2678
MAXMIN, RZFBF	1273.7203	1042.1833
MAXMIN, ZFBF	1267.2368	1037.2915

form. For PFS, we have $g(\mathcal{R}_1,\ldots,\mathcal{R}_Q)=\sum_{q=1}^{Q}\log(\mathcal{R}_q)$, yielding the solution $\nu_q=\frac{1}{Q}$ for all q. For max-min fairness, we have $g(\mathcal{R}_1,\ldots,\mathcal{R}_Q)=\min_q \mathcal{R}_q$, yielding the solution $\nu_q=\frac{\frac{1}{\mathcal{R}_q^*}}{\sum_q \frac{1}{\mathcal{R}_q^*}}$.

The spectral efficiency $\mathcal{R}_q^*$ can be optimized independently for each pattern $\mathcal{A}_q$. For a given JSDM precoding scheme, we need to search over the number of downlink streams in each group. This is a multi-dimensional integer search over the parameters $\{S_{l,g}\}$ for all groups $(l,g)\in\mathcal{A}_q$. In addition, we should optimize with respect to the power allocation to the downlink data streams, as mentioned before. In order to obtain a tractable problem, we resort to good heuristics. Following the design guideline given in Remark 8, we know that the ratio $S_{l,g}/b_{l,g}$ should be approximately the same for the optimal $S_{l,g}$ for all groups (l,g) with similar geometry, i.e., belonging to the same region. Hence, we fix this ratio to be the same for all groups in the same region, and indicate it as α_l. In addition, as done before, we restrict to equal power allocation to all the downlink streams. Indicating this common per-stream power value by $\bar{P}$, and letting $R_{l,g,k}(\bar{P})$ denote the rate of the k-th stream of group (l,g) as a function of $\bar{P}$, calculated according to the methods given in Section V and Appendix A, for given MU-MIMO precoding scheme, the optimization with respect to $\{\alpha_l\}$ is expressed by

$$\begin{aligned} \max \quad & \sum_{l\in\mathcal{A}_q}\sum_{g=1}^{G_l}\sum_{k=1}^{S_{l,g}} R_{l,g,k}(\bar{P}) \\ \text{subject to} \quad & S_{l,g}=\lfloor \alpha_l b_{l,g}\rfloor \\ & \bar{P}=\frac{P}{\sum_{l\in\mathcal{A}_q}\sum_{g=1}^{G_l} S_{l,g}}. \end{aligned} \tag{80}$$

Notice that for a pattern with $|\mathcal{A}_q|$ regions, (80) consists of a $|\mathcal{A}_q|$-dimensional search over the real parameters $\alpha_l\in[0,1]$, which is tractable when $|\mathcal{A}_q|$ is small (in our case, $|\mathcal{A}_q|=2$).

Fig. 10 shows the sum spectral efficiency $\sum_{g=1}^{G_l}\sum_{k=1}^{S_{l,g}} R_{l,g,k}(\bar{P})$ for each annular region $l=1,\ldots,8$ in the setup of Fig. 9 with system parameters given at the beginning of this section, resulting from the above optimization for both DFT pre-beamforming and approximate BD using PGP with RZFBF and ZFBF precoding. The corresponding sum spectral efficiencies under PFS and max-min fairness scheduling are reported in Table I.

IX. Concluding remarks

In this work we proposed Joint Space-Division and Multiplexing (JSDM), a novel approach to MU-MIMO downlink that requires reduced channel estimation downlink training overhead and CSIT feedback and therefore is potentially suited to FDD systems, despite using a large number of BS antennas. JSDM exploits the fact that for large BSs, mounted on the top of a building or on a dedicated tower, channel vectors are far from isotropically distributed. Instead, their dominant eigenspace has dimension much smaller than the number of BS antennas. Different groups of users are selected, such that the users in each group share (approximately) the same dominant eigenspace, and the eigenspaces of different groups are nearly orthogonal. JSDM serves simultaneously such groups of users, and multiple users in each group. The separation of the groups in the spatial domain (space-division) is obtained through a pre-beamforming matrix that depends only on the channel covariance matrices, while the multiplexing of multiple users in each group is obtained via linear MU-MIMO precoding based on the instantaneous "effective" channel, including the pre-beamforming. It turns out that the effective channel has reduced dimensionality with respect to the original multi-antenna multiuser channel, especially with a "per-group processing" (PGP) approach, i.e., where each group is individually pre-coded, disregarding the inter-group interference. JSDM with PGP can be regarded as a generalization of sectorization, where each group acts as a directional sector, and in each sector we apply MU-MIMO spatial multiplexing, disregarding inter-sector interference.

We showed that when the collection of the channel covariance eigenvectors of the groups forms a tall unitary matrix, then JSDM with PGP is optimal, in the sense that it can achieve the capacity of the underlying MU-MIMO channel with full instantaneous CSIT. Then, using Szego's asymptotic theory of large Toeplitz random matrices, we showed that when the BS is equipped with a large linear uniform array, this tall unitary condition is closely approached, and the pre-beamforming matrix can be obtained by selecting an appropriate subset of columns of a unitary DFT matrix. In fact, under these assumptions the accurate estimation of the channel covariance matrix is not needed, and just a coarse estimation of the AoA range for each group is sufficient, as long as the AoA ranges of different groups do not overlap in the azimuth angle domain. Finally, we extended our approach to the case of 3D beamforming, considering rectangular arrays and pre-beamforming in the elevation angle (vertical) direction. In this case, the proposed JSDM scheme partitions the cell into concentric annular regions, and serves groups of users with different azimuth angle in each region. We demonstrated the effectiveness of the proposed scheme in the case of a typical cell size, typical propagation pathloss, and a large rectangular antenna array mounted on the face of a tall building. In our case, under ideal CSIT, unprecedented spectral efficiencies of the order of 1000 bit/s/Hz per sector are achieved under various fairness criteria and pre-beamforming techniques.

We also considered the problem of downlink channel estimation and provided formulas for the asymptotic "deterministic equivalent" approximation of the achievable receiver SINR, which allows efficient calculation of the system

performance without resorting to lengthy Monte Carlo simulation. For a realistic SNR range around 20 dB, the effect of noisy CSIT can be quantified in $\approx 30\%$ loss with respect to the ideal CSIT case. Hence, spectral efficiencies of ≈ 700 bit/s/Hz can be expected for the massive 3D JSDM system scenario.

The design of a JSDM system involves many choices: effective rank $r_g^\star$ of the channel covariance matrix for each group, pre-beamforming dimension b_g, number of users (downlink streams) for each group S_g, for given pre-beamforming design, operating SNR, and MU-MIMO precoding scheme. In the case of 3D beamforming, this optimization is significantly more complicated since it has to be repeated for groups of annular regions served simultaneously by the vertical beamforming. One of the main merits of this paper is to provide simple and solid design criteria for such a system, based on the insight gained by the asymptotic analysis. In fact, a brute-force search over the whole parameter space becomes quickly infeasible for practical system scenarios.

We conclude this work by pointing out two interesting related topics, which are left for future work: 1) user group formation; 2) estimation of the channel covariance matrix dominant eigenspace. User group formation considers clustering algorithms that, given K users each of which is characterized by its channel covariance dominant eigenspace, forms groups of users that can be served simultaneously using JSDM, such that the system spectral efficiency is maximized. In order to enable user group formation and JSDM, the dominant eigenspace of each user must be estimated from noisy samples of the received signal. Here, the problem is that for a large number of BS antennas the channel covariance matrix is high-dimensional, and the dimension is typically comparable with the number of samples. Hence, the common wisdom on "sample covariance" estimation does not apply, and more sophisticated techniques must be used (e.g., [19, Ch. 17], [14], [16]).

APPENDIX A
DETERMINISTIC EQUIVALENTS FOR THE SINR WITH PGP AND NOISY CSIT

We provide fixed-point equations for the calculation of the deterministic equivalent approximations of the SINR for JSDM with PGP, noisy CSIT and the two types of linear precoding considered in this paper, namely, RZFBF and ZFBF. Notice that these expressions hold for arbitrary pre-beamforming matrices, as long as they are fixed constants independent of the instantaneous channel matrix realizations. In particular, they hold for (approximated) BD and DFT pre-beamforming. We consider the general case of group parameters $\{S_g\}$, $\{b_g\}$, with equal power per stream, $P_{g_k} = \frac{P}{S}$ for all g_k. The formulas below are a direct application of the results in [10]. Their derivation is lengthy but somehow straightforward after realizing that all the assumption in [10] apply to our case. In the spirit of striking a good balance between usefulness, conciseness and completeness, we report the formulas without the details of their derivation.

A. Regularized Zero Forcing Precoding

For users in group g, the regularized zero forcing precoding matrix is given by

$$\boldsymbol{P}_{g,\mathrm{rzf}} = \bar{\zeta}_g(\widehat{\mathbf{H}}_g\widehat{\mathbf{H}}_g^{\mathsf{H}} + b_g\alpha\boldsymbol{I}_{b_g})^{-1}\widehat{\mathbf{H}}_g, \tag{81}$$

where $\widehat{\mathbf{H}}_g$ is the matrix formed by the channel estimates $\widehat{\mathbf{h}}_{g_k}$ obtained as in (50). The power normalization factor $\bar{\zeta}_g$ is given by

$$\bar{\zeta}_g^2 = \frac{S_g}{\mathrm{tr}\left(\boldsymbol{P}_{g,\mathrm{rzf}}^{\mathsf{H}}\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g\boldsymbol{P}_{g,\mathrm{rzf}}\right)} \tag{82}$$

Letting $\widehat{\bar{\boldsymbol{K}}}_g = (\widehat{\mathbf{H}}_g\widehat{\mathbf{H}}_g^{\mathsf{H}} + b_g\alpha\boldsymbol{I}_{b_g})^{-1}$, the SINR of user g_k is given by

$$\widehat{\gamma}_{g_k,\mathrm{pgp,icsi}} = \frac{\frac{P}{S}\bar{\zeta}_g^2|\widehat{\mathbf{h}}_{g_k}^{\mathsf{H}}\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_k}|^2}{\frac{P}{S}\bar{\zeta}_g^2|\widehat{\mathbf{e}}_{g_k}^{\mathsf{H}}\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_k}|^2 + \sum_{j\neq k}\frac{P}{S}\bar{\zeta}_g^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_g\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_j}|^2 + \sum_{g'\neq g,j}\frac{P}{S}\bar{\zeta}_{g'}^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_{g'}\widehat{\bar{\boldsymbol{K}}}_{g'}\widehat{\mathbf{h}}_{g'_j}|^2 + 1} \tag{83}$$

where "csi" denotes noisy CSIT. The deterministic equivalent of the SINR in this case is given by

$$\widehat{\gamma}_{g_k,\mathrm{pgp,rzf,csi}} - \widehat{\gamma}^{o}_{g_k,\mathrm{pgp,rzf,csi}} \overset{M\to\infty}{\longrightarrow} 0 \tag{84}$$

with

$$\widehat{\gamma}^{o}_{g_k,\mathrm{pgp,rzf,csi}} = \frac{\frac{P}{S}\widehat{\bar{\zeta}}_g^2(\widehat{\bar{m}}_g^o)^2}{\frac{P}{S}\widehat{\bar{\zeta}}_g^2\widehat{\bar{E}}_g^o + \frac{P}{S}\widehat{\bar{\zeta}}_g^2\widehat{\Upsilon}_{g,g}^o + \left(1+\sum_{g'\neq g}\frac{P}{S}\widehat{\bar{\zeta}}_{g'}^2\widehat{\Upsilon}_{g,g'}^o\right)\left(1+\widehat{\bar{m}}_g^o\right)^2} \tag{85}$$

where $\widehat{\bar{\zeta}}_g^2 = \frac{1}{\widehat{\bar{\Gamma}}_g^o}$ and the quantities $\widehat{\bar{m}}_g^o$, $\widehat{\bar{\Upsilon}}_{g,g}^o$, $\widehat{\bar{\Upsilon}}_{g,g'}^o$ and $\widehat{\bar{\Gamma}}_g^o$ are given by

$$\widehat{\bar{m}}_g^o = \frac{1}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right) \tag{86}$$

$$\widehat{\bar{\boldsymbol{T}}}_g = \left(\frac{S_g}{b_g}\frac{\widehat{\bar{\boldsymbol{R}}}_g}{1+\widehat{\bar{m}}_g^o} + \alpha\boldsymbol{I}_{b_g}\right)^{-1} \tag{87}$$

$$\widehat{\bar{\Gamma}}_g^o = \frac{1}{b_g}\frac{\widehat{\bar{n}}_g}{(1+\widehat{\bar{m}}_g^o)^2} \tag{88}$$

$$\widehat{\bar{n}}_g = \frac{\frac{1}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(1+\widehat{\bar{m}}_g^o)^2}} \tag{89}$$

$$\widehat{\bar{E}}_g^o = \frac{1}{b_g}\frac{\frac{1}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g(\bar{\boldsymbol{R}}_g-\widehat{\bar{\boldsymbol{R}}}_g)\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(1+\widehat{\bar{m}}_g^o)^2}} \tag{90}$$

$$\widehat{\bar{\Upsilon}}_{g,g}^o = (1+\widehat{\bar{m}}_g^o)^2A_1 - \left[2\widehat{\bar{m}}_g^o(1+\widehat{\bar{m}}_g^o) - (\widehat{\bar{m}}_g^o)^2\right]A_2 \tag{91}$$

$$A_1 = \frac{1}{b_g}(S_g-1)\frac{\widehat{\bar{n}}_{g,g,1}}{(1+\widehat{\bar{m}}_g^o)^2} \tag{92}$$

$$A_2 = \frac{1}{b_g}(S_g-1)\frac{\widehat{\bar{n}}_{g,g,2}}{(1+\widehat{\bar{m}}_g^o)^2} \tag{93}$$

$$\widehat{\bar{n}}_{g,g,1} = \frac{\frac{1}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\bar{\boldsymbol{R}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(1+\widehat{\bar{m}}_g^o)^2}} \tag{94}$$

$$\widehat{\bar{n}}_{g,g,2} = \frac{\frac{1}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(1+\widehat{\bar{m}}_g^o)^2}} \tag{95}$$

$$\widehat{\bar{\Upsilon}}_{g,g'}^o = \frac{S_{g'}}{b_{g'}}\frac{\widehat{\bar{n}}_{g',g}}{(1+\widehat{\bar{m}}_{g'}^o)^2} \tag{96}$$

$$\widehat{\bar{n}}_{g',g} = \frac{\frac{1}{b_{g'}}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\boldsymbol{B}_{g'}^{\mathsf{H}}\boldsymbol{R}_g\boldsymbol{B}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\right)}{1-\frac{\frac{S_{g'}}{b_{g'}}\text{tr}\left(\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\right)}{b_{g'}(1+\widehat{\bar{m}}_{g'}^o)^2}} \tag{97}$$

B. Zero Forcing Precoding

For $\alpha = 0$, the precoding matrix in (81) reduces to the zero forcing precoding matrix given by

$$\boldsymbol{P}_{g,\text{zf}} = \bar{\zeta}_g\widehat{\mathbf{H}}_g(\widehat{\mathbf{H}}_g^{\mathsf{H}}\widehat{\mathbf{H}}_g)^{-1} \tag{98}$$

where $\bar{\zeta}_g$ is the power normalization factor given by

$$\bar{\zeta}_g^2 = \frac{S_g}{\text{tr}(\boldsymbol{P}_{g,\text{zf}}^{\mathsf{H}}\boldsymbol{B}_g\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{P}_{g,\text{zf}})} \tag{99}$$

Letting $\widehat{\bar{\boldsymbol{K}}}_g = \widehat{\mathbf{H}}_g(\widehat{\mathbf{H}}_g^{\mathsf{H}}\widehat{\mathbf{H}}_g)^{-2}\widehat{\mathbf{H}}_g^{\mathsf{H}}$, the SINR of user g_k is given by

$$\widehat{\gamma}_{g_k,\mathrm{pgp,zf,csi}} = \frac{\frac{P}{S}\widehat{\bar{\zeta}}_g^2|\widehat{\mathbf{h}}_{g_k}^{\mathsf{H}}\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_k}|^2}{\frac{P}{S}\widehat{\bar{\zeta}}_g^2|\widehat{\mathbf{e}}_{g_k}^{\mathsf{H}}\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_k}|^2 + \sum_{j\neq k}\frac{P}{S}\widehat{\bar{\zeta}}_g^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_g\widehat{\bar{\boldsymbol{K}}}_g\widehat{\mathbf{h}}_{g_j}|^2 + \sum_{g'\neq g,j}\frac{P}{S}\widehat{\bar{\zeta}}_{g'}^2|\boldsymbol{h}_{g_k}^{\mathsf{H}}\boldsymbol{B}_{g'}\widehat{\bar{\boldsymbol{K}}}_{g'}\widehat{\mathbf{h}}_{g_j'}|^2 + 1} \tag{100}$$

The deterministic equivalent of the SINR is given as

$$\widehat{\gamma}_{g_k,\mathrm{pgp,zf,csi}} - \widehat{\gamma}^{o}_{g_k,\mathrm{pgp,zf,csi}} \overset{M\to\infty}{\longrightarrow} 0 \tag{101}$$

where $\widehat{\gamma}^{o}_{g_k,\mathrm{pgp,zf,csi}}$ is given by

$$\begin{aligned}\widehat{\gamma}^{o}_{g_k,\mathrm{pgp,zf,csi}} &= \frac{\frac{P}{S}\widehat{\bar{\zeta}}_g^2}{1 + \frac{P}{S}\widehat{\bar{\zeta}}_g^2\frac{\widehat{\bar{E}}_g^o}{(\widehat{\bar{m}}_g^o)^2} + \frac{P}{S}\widehat{\bar{\zeta}}_g^2\widehat{\Upsilon}^o_{g,g} + \sum_{g'\neq g}\frac{P}{S}\widehat{\bar{\zeta}}_{g'}^2\widehat{\Upsilon}^o_{g,g'}} \\ &= \frac{\frac{P}{S}\widehat{\bar{\zeta}}_g^2}{1 + \frac{P}{S}\widehat{\bar{\zeta}}_g^2 S_g\frac{\widehat{\bar{E}}_g^o}{(\widehat{\bar{m}}_g^o)^2} + \sum_{g'\neq g}\frac{P}{S}\widehat{\bar{\zeta}}_{g'}^2\widehat{\Upsilon}^o_{g,g'}}\end{aligned} \tag{102}$$

with $\widehat{\bar{\zeta}}_g^2 = \frac{1}{\widehat{\bar{\Gamma}}_g^o}$ and the quantities $\widehat{\bar{\Gamma}}_g^o$, $\widehat{\bar{\Upsilon}}_{g,g'}^o$, and $\widehat{\bar{m}}_g^o$ are given by [12]

$$\widehat{\bar{m}}_g^o = \frac{1}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right) \tag{103}$$

$$\widehat{\bar{\boldsymbol{T}}}_g = \left(\frac{S_g}{b_g}\frac{\widehat{\bar{\boldsymbol{R}}}_g}{\widehat{\bar{m}}_g^o} + \boldsymbol{I}_{b_g}\right)^{-1} \tag{104}$$

$$\widehat{\bar{\Gamma}}_g^o = \frac{1}{b_g}\frac{\widehat{\bar{n}}_g}{(\widehat{\bar{m}}_g^o)^2} \tag{105}$$

$$\widehat{\bar{\Upsilon}}_{g,g'}^o = \frac{S_{g'}}{b_{g'}}\frac{\widehat{\bar{n}}_{g',g}}{(\widehat{\bar{m}}_{g'}^o)^2} \tag{106}$$

$$\widehat{\bar{n}}_g = \frac{\frac{1}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(\widehat{\bar{m}}_g^o)^2}} \tag{107}$$

$$\widehat{\bar{n}}_{g',g} = \frac{\frac{1}{b_{g'}}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\boldsymbol{B}_{g'}^{\mathsf{H}}\boldsymbol{R}_g\boldsymbol{B}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\right)}{1-\frac{\frac{S_{g'}}{b_{g'}}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\widehat{\bar{\boldsymbol{R}}}_{g'}\widehat{\bar{\boldsymbol{T}}}_{g'}\right)}{b_{g'}(\widehat{\bar{m}}_{g'}^o)^2}} \tag{108}$$

$$\widehat{\bar{\Upsilon}}_{g,g}^o = A_1 - A_2 \tag{109}$$

$$A_1 = \frac{1}{b_g}(S_g-1)\frac{\widehat{\bar{n}}_{g,g,1}}{(\widehat{\bar{m}}_g^o)^2} \tag{110}$$

$$A_2 = \frac{1}{b_g}(S_g-1)\frac{\widehat{\bar{n}}_{g,g,2}}{(\widehat{\bar{m}}_g^o)^2} \tag{111}$$

$$\widehat{\bar{n}}_{g,g,1} = \frac{\frac{1}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\bar{\boldsymbol{R}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(\widehat{\bar{m}}_g^o)^2}} \tag{112}$$

$$\widehat{\bar{n}}_{g,g,2} = \frac{\frac{1}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(\widehat{\bar{m}}_g^o)^2}} \tag{113}$$

$$\widehat{\bar{E}}_g^o = \frac{1}{b_g}\frac{\frac{1}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g(\bar{\boldsymbol{R}}_g-\widehat{\bar{\boldsymbol{R}}}_g)\widehat{\bar{\boldsymbol{T}}}_g\right)}{1-\frac{\frac{S_g}{b_g}\mathrm{tr}\left(\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\widehat{\bar{\boldsymbol{R}}}_g\widehat{\bar{\boldsymbol{T}}}_g\right)}{b_g(\widehat{\bar{m}}_g^o)^2}} \tag{114}$$

In order to obtain the desired expression (102), we notice that

$$\frac{\widehat{\bar{E}}_g^o}{(\widehat{\bar{m}}_g^o)^2} + \widehat{\bar{\Upsilon}}_{g,g}^o = S_g\frac{\widehat{\bar{E}}_g^o}{(\widehat{\bar{m}}_g^o)^2} \tag{115}$$

[12] It is easy to see that when $\boldsymbol{B}_g^{\mathsf{H}}\boldsymbol{B}_g = \boldsymbol{I}_{b_g}$, $\widehat{\bar{n}}_g = \widehat{\bar{m}}_g$

APPENDIX B
GENERAL FORMULA FOR $S(\xi)$

We find the general expression for $S(\xi)$ defined in (56) for $r_m = [\boldsymbol{R}]_{\ell,\ell-m}$ with $[\boldsymbol{R}]_{m,p}$ given by (55), without any restriction on the AoA range. We have:

$$\begin{aligned} S(\xi) &= \sum_{m=-\infty}^{\infty} \left[\frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} e^{-j2\pi D m \sin(\alpha)} d\alpha \right] e^{-j2\pi\xi m} \\ &= \frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} \left[\sum_{m=-\infty}^{\infty} e^{-j2\pi m(D\sin(\alpha)+\xi)} \right] d\alpha \\ &= \frac{1}{2\Delta} \int_{-\Delta+\theta}^{\Delta+\theta} \left[\sum_{m=-\infty}^{\infty} \delta(D\sin(\alpha)+\xi-m) \right] d\alpha \\ &= \frac{1}{2\Delta} \int \left[\sum_{m=-\infty}^{\infty} \delta(z+\xi-m) \right] \frac{dz}{\sqrt{D^2-z^2}}, \end{aligned} \tag{116}$$

The limits in (116) depend on the range of $[\theta-\Delta, \theta+\Delta]$. We distinguish the following cases:

1) For $\theta+\Delta < -\frac{\pi}{2}, \theta-\Delta > \frac{\pi}{2}$ and $-\frac{\pi}{2} \le \theta-\Delta < \theta+\Delta \le \frac{\pi}{2}$, (116) becomes

$$\sum_{m=-\infty}^{\infty} \left[\int_{\min(D\sin(\theta-\Delta), D\sin(\theta+\Delta))}^{\max(D\sin(\theta-\Delta), D\sin(\theta+\Delta))} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right] \tag{117}$$

2) For $\theta-\Delta < -\frac{\pi}{2}, \theta+\Delta > \frac{\pi}{2}$, (116) becomes

$$\sum_{m=-\infty}^{\infty} \left[\int_{-D}^{D\sin(\theta-\Delta)} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} + \int_{-D}^{D} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right. \tag{118}$$

$$\left. + \int_{D\sin(\theta+\Delta)}^{D} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right] \tag{119}$$

3) For $\theta-\Delta < -\frac{\pi}{2}, -\frac{\pi}{2} \le \theta+\Delta \le \frac{\pi}{2}$, (116) becomes

$$\sum_{m=-\infty}^{\infty} \left[\int_{-D}^{D\sin(\theta-\Delta)} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} + \int_{-D}^{D\sin(\theta+\Delta)} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right] \tag{120}$$

4) For $-\frac{\pi}{2} \le \theta-\Delta \le \frac{\pi}{2}, \theta+\Delta > \frac{\pi}{2}$, (116) becomes

$$\sum_{m=-\infty}^{\infty} \left[\int_{D\sin(\theta-\Delta)}^{D} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} + \int_{D\sin(\theta+\Delta)}^{D} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right] \tag{121}$$

Now, owing to the property of the Dirac delta function, we have

$$\sum_{m=-\infty}^{\infty} \left[\int_{A}^{B} \delta(z+\xi-m) \frac{dz}{\sqrt{D^2-z^2}} \right] = \sum_{m\in[A+\xi, B+\xi]} \frac{1}{\sqrt{D^2-(m-\xi)^2}} \tag{122}$$

as a result of which we can write $S(\xi)$ for the cases identified above as

1) Case $\theta+\Delta < -\frac{\pi}{2}, \theta-\Delta > \frac{\pi}{2}$ and $-\frac{\pi}{2} \le \theta-\Delta < \theta+\Delta \le \frac{\pi}{2}$

$$S(\xi) = \frac{1}{2\Delta} \sum_{m\in[\min(D\sin(-\Delta+\theta), D\sin(\Delta+\theta))+\xi, \max(D\sin(-\Delta+\theta), D\sin(\Delta+\theta))+\xi]} \frac{1}{\sqrt{D^2-(m-\xi)^2}} \tag{123}$$

2) Case $\theta-\Delta<-\frac{\pi}{2},\theta+\Delta>\frac{\pi}{2}$

$$\begin{aligned} S(\xi) &= \frac{1}{2\Delta}\sum_{m\in[-D+\xi,D\sin(-\Delta+\theta)+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}}+\frac{1}{2\Delta}\sum_{m\in(-D+\xi,D+\xi)}\frac{1}{\sqrt{D^2-(m-\xi)^2}} \\ &\quad +\frac{1}{2\Delta}\sum_{m\in[D\sin(\Delta+\theta)+\xi,D+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}} \end{aligned}$$

3) Case $\theta-\Delta<-\frac{\pi}{2},-\frac{\pi}{2}\leq\theta+\Delta\leq\frac{\pi}{2}$

$$S(\xi)=\frac{1}{2\Delta}\sum_{m\in[-D+\xi,D\sin(-\Delta+\theta)+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}}+\frac{1}{2\Delta}\sum_{m\in(-D+\xi,D\sin(\Delta+\theta)+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}} \quad (124)$$

4) Case $-\frac{\pi}{2}\leq\theta-\Delta\leq\frac{\pi}{2},\theta+\Delta>\frac{\pi}{2}$

$$S(\xi)=\frac{1}{2\Delta}\sum_{m\in[D\sin(-\Delta+\theta)+\xi,D+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}}+\frac{1}{2\Delta}\sum_{m\in[D\sin(\Delta+\theta)+\xi,D+\xi)}\frac{1}{\sqrt{D^2-(m-\xi)^2}} \quad (125)$$

It is easy to see that the formula reduces to (62) when $-\frac{\pi}{2}\leq\theta-\Delta<\theta+\Delta\leq\frac{\pi}{2}$. Taking the limits from $-\pi$ to π recovers the Fourier transform of the Bessel J_0 function commonly used to model correlated Rayleigh fading in an isotropic scattering environment [34], given by $\frac{1}{\pi}\frac{\mathrm{rect}(\xi/2D)}{\sqrt{D^2-\xi^2}},\xi\in[-1/2,1/2]$ for $D\in[0,\frac{1}{2}]$.

REFERENCES

[1] L. Zheng and D. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Inform. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. on Inform. Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[3] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser mimo downlink," *IEEE Trans. on Commun.*, vol. 59, no. 8, 2011.

[4] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling," *IEEE Trans. on Inform. Theory*, vol. 58, no. 5, pp. 2911 – 2934, 2012.

[5] T. Gou, C. Wang, and S. Jafar, "Aiming perfectly in the dark-blind interference alignment through staggered antenna switching," in *IEEE Global Telecommunications Conference (GLOBECOM 2010)*. IEEE, 2010, pp. 1–5.

[6] M. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 1188–1195.

[7] M. Kobayashi and G. Caire, "On the net dof comparison between zf and mat over time-varying miso broadcast channels," in *IEEE International Symposium on Information Theory*, 2012.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas," *IEEE Trans. on Wireless Commun.*, vol. PP, no. 99, pp. 1 – 14, 2012.

[10] R. Couillet, S. Wagner, and M. Debbah, "Asymptotic analysis of linear precoding techniques in correlated multi-antenna broadcast channels," *CoRR*, vol. abs/0906.3682, 2009. [Online]. Available: http://arxiv.org/abs/0906.3682

[11] H. Clayton Shepard, N. Anand, L. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," *Rice University Technical Report*, 2012. [Online]. Available: http://networks.rice.edu/papers/argos.pdf

[12] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Commun.*, vol. 48, no. 3, pp. 502–513, 2000.

[13] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case," *IEEE Trans. on Inform. Theory*, vol. 58, no. 2, pp. 1043–1068, 2012.

[14] T. Marzetta, G. Tucci, and S. Simon, "A random matrix–theoretic approach to handling singular covariance estimates," *IEEE Trans. on Inform. Theory*, vol. 57, no. 9, 2011.

[15] B. Hochwald and T. Maretta, "Adapting a downlink array from uplink measurements," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 3, pp. 642–653, 2001.

[16] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. on Inform. Theory*, vol. 54, no. 11, pp. 5113–5129, 2008.

[17] U. Grenander and G. Szegő, *Toeplitz forms and their applications*. Chelsea Pub Co, 1984.

[18] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-scale Multiple-antenna Systems," *CoRR*, vol. abs/1203.5924, 2012. [Online]. Available: http://arxiv.org/abs/1203.5924

[19] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge Univ Pr, 2011.

[20] A. Eriksson, P. Stoica, and T. Soderstrom, "On-line subspace algorithms for tracking moving sources," *IEEE Trans. on Sig. Proc.*, vol. 42, no. 9, pp. 2319–2330, 1994.

[21] B. Hochwald and T. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 3, pp. 642–653, 2001.

[22] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[23] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *Information Theory, IEEE Transactions on*, vol. 52, no. 9, pp. 3936–3964, 2006.

[24] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 2, pp. 461–471, 2004.

[25] P. Viswanath, N. David, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, p. 1277, 2002.

[26] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[27] T. Al-Naffouri, M. Sharif, and B. Hassibi, "How much does transmit correlation affect the sum-rate scaling of mimo gaussian broadcast channels?" *IEEE Trans. on Commun.*, vol. 57, no. 2, pp. 562–572, 2009.

[28] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channels with partial side information," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, pp. 506–522, 2005.

[29] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Trans. on Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.

[30] W. Yu, "Sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," *IEEE Trans. on Inform. Theory*, vol. 52, no. 2, pp. 754–759, 2006.

[31] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in UL/DL of cellular networks: How many antennas do we need?" *CoRR*, vol. abs/1107.1709, 2011. [Online]. Available: http://arxiv.org/abs/1107.1709

[32] R. Gray, *Toeplitz and circulant matrices: A review*. Now Pub, 2006.

[33] A. Lapidoth, *A Foundation in Digital Communications*. Cambridge University Press, 2009.

[34] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Transactions on Communications Systems*, vol. 11, no. 4, pp. 360–393, 1963.

[35] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[36] S. Ramprashad, G. Caire, and H. Papadopoulos, "Cellular and network mimo architectures: Mu-mimo spectral efficiency and costs of channel state information," in *Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 1811–1818.

[37] ——, "A joint scheduling and cell clustering scheme for mu-mimo downlink with limited coordination," in *IEEE International Conference on Communications (ICC)*. IEEE, 2010, pp. 1–6.

[38] G. Caire, S. Ramprashad, and H. Papadopoulos, "Rethinking network mimo: Cost of csit, performance analysis, and architecture comparisons," in *Information Theory and Applications Workshop (ITA), 2010*. IEEE, 2010, pp. 1–10.